Three essays on poverty measurement and risk protection

Sandoval Moreno, Carlos

DOI:
10.26116/center-lis-1938

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Three Essays on Poverty Measurement and Risk Protection

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. K. Sijtsma, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Portrettenzaal van de Universiteit op

maandag 16 december 2019 om 10.00 uur

door

Carlos Eduardo Sandoval Moreno

geboren op 19 december 1982 te Bogotá, Colombia.

Promotor:
Prof. dr. Arthur van Soest
Copromotor:
dr. Martin Salm

Promotiecommissie:
Prof. dr. Bertrand Melenberg
Prof. dr. Tobias Klein
Prof. dr. Chris Elbers
dr. João Pedro Azevedo

Three essays on poverty measurement and risk protection

By

Carlos Eduardo Sandoval

Acknowledgements

This thesis is the result of almost four years of work and the moral, financial and intellectual support of various persons that have been part of my life at different stages.

My mother and family support me through all of this process, with their motivation and financial support when needed. My partner, Irina España shared with me the ups and downs of doing a PhD.

During the process of writing this dissertation I frequently engaged in valuable and long discussion with Laura Capera and Santiago Bohórquez related to the topics considered here. Richard Jaimes offered me top-notch advice on the methods used in chapter 3 and Victor Gonzalez was an unexpected friend in moments of doubt.

My advisors Arthur van Soest and Martin Salm were the academic guidance throughout this long and sometimes discouraging process. Arthur provided me with direction at all levels: conceptual, methodological, and sometimes even editing advice which is exceptional for a supervisor. Martin helped me with the development of the fourth chapter, but also gave me valuable comments on the entire manuscript. He offered me detailed recommendations along with high academic requirements for the completion of that chapter.

I want to thank specially the members of my committee Prof. Tobias Klein, Prof. Bertrand Melenberg, Prof. Chris Elbers and Dr. João Pedro Azevedo for taking the time of reading and commenting in detail this dissertation. Additionally, participants of the Seminario de Economía of Banco de la República in Bogotá and Medellín and participants on the Workshop on Poverty and Inequality at the 2019 Ridge-Lacea forum provided useful insights in various chapters of this dissertation.

Finally, Banco de la República of Colombia and COLFUTURO offered me partial financial support at various stages in the process of completing this PhD.

Table of contents

Chapter 1. Preface

After working most of my career in poverty and development topics the question on what are the underlying structural causes of poverty always keep revolving around my mind. It is shocking to observe first hand how some households struggle to keep living just out of poverty, while others never manage to get out of it. This is specially concerning in Latin America and the Caribbean region which is well known for being the most unequal region in the word.

This motivation fuelled by curiosity and a certain knowledge on the field push me to research deeper in poverty measurement topics and how shocks may affect permanently the poverty status of households. Indeed, an overview on the poverty measurement literature makes evident that they are very concerned on doing descriptive figures on the evolution of poverty across time but not that much in which are the structural causes of this poverty (Foster, Greer, Thorbecke, 1984; Foster, 2009; Levy, 1977). For being able to find studies that try to understand the determinants on income and how they evolve though time you have to look in all the strand of intergenerational mobility literature (Becker and Tomen, 1979; Galor and Moav, 2004; Heckman and Mosso, 2014), poverty trap literature (Banerjee and Newman, 1993; Galor and Zeira, 1993; Dasgupta and Ray, 1986) and in general on the studies concerning on the structural modelling of income (Carpio, Wohlgenant and Safley, 2008). Nevertheless, this literature leaves untouched the topic on how to measure poverty founded in a structural model for the household income.

The detachment between a strand of literature centred around the description of evolution of poverty and another strand interested in the structural modelling of income worry me not only as an economist, but also as a policy maker, as this disconnection may end up in a wasted opportunity to design more efficient and effective anti poverty policies.

Nonetheless, as is common in research, I was not the first one considering the lack of economic modelling a problem in poverty measurement. Certainly, Carter and Barrett (2006) in their seminal work pioneered the idea that a poverty measurement should include also some form of behavioural foundation. In their work they give the basis for a poverty measure that has some basic modelling of the behaviour of economic agents. This poverty measure is in turn based on the concept of poverty traps derived from macroeconomics. An abundant set of studies have followed this theoretical framework trying to find the existence of poverty traps in different socio-economic contexts (see for instance Dutta, 2015; You, 2014 and Giesbert and Schiendler, 2012 among others). However this new ideas also comes with shortcomings. First, Carter and Barrett (2006) model is based in one very particular behavioural result. Second, they ignore in their model the effect that income shocks may have in a poverty trap. Third, all the works based on this literature focus almost exclusively in rural areas. Fourth, surprisingly none of this empirical literature estimate which percentage of households are trapped in poverty according to the Carter and Barrett (2006) model, limiting themselves to test only the existence of a poverty trap.

But why it is relevant to quantify and identify households trapped in poverty, and not only to identify the existence of a poverty trap?. Carter and Barrett (2006) model proposes to analyse the dynamic evolution of productive assets in possession of the household. This is operationalized analysing the dynamic of the part of the income explained by household productive assets. If the dynamic evolution of the productive assets is not enough to produce income above a given monetary poverty line, the household can be considered trapped in dynamic structural poverty. In other words, if the productive assets in possession of the household are not enough to produce income above the monetary poverty line, and the productive asset accumulation process is not sufficient to produce income above the poverty line in subsequent periods, the household is trapped in poverty and have no hopes to escape it.

A poverty measure based on this idea would allow to differentiate which households need a deep, long run intervention that supports them in the productive asset accumulation process to leave poverty, from those households in a transitory monetary poverty situation that may be alleviated trough temporarily cash transfers.

This thesis aims to add on these ideas improving in various fronts and intends to create a bridge among the two streams of literature previously described which I consider inexplicable disconnected.

The second chapter “Welfare and poverty traps in urban Colombia: an approach from the asset dynamics perspective”, estimates the Carter an Barrett (2006) model for urban areas in Colombia. This is challenging because estimating the part of the income that depends on productive assets in urban areas brings some technical complications not present in rural areas. First, in rural areas and developing economies with low social mobility, it is reasonable to assume that most household members work in their own farm, making the productive assets in possession of the household a good set of explanatory variables for household income. Second, as a consequence of the first point, household surveys in rural areas ask for detailed information on household productive assets. None of those conditions hold in urban areas, as household members may be employees or self-employed, and information in productive assets is not very detailed.

After proposing various methods to overcome these technical difficulties and applying the Carter and Barrett (2006) model I found that under the particular behavioural assumptions proposed by Carter and Barrett (2006) there is not a poverty trap in urban areas in Colombia.

The third chapter titled "Household income dynamics and income convergence: a new way to measure poverty", is still based on the idea of measuring the existence of a poverty trap modelling the evolution of the part of the income explained by productive assets. Nevertheless, this chapter proposes various improvements on the Carter and Barrett (2006) model. First, it explores a different functional form of the asset dynamics to test the existence of a structural dynamic poverty trap following Phillips and Sul (2007). Second, this methodology allows to identify and quantify what households are trapped in dynamic structural poverty. Under this estimation methodology I found that 5.8% of households in three main urban areas in Colombia are trapped in dynamic structural poverty. Third, the methodology proposed here addresses the fact that, despite the asset dynamics being the most important component on income evolution, shocks to income cannot be ignored and must be modelled jointly. On this respect, I propose to complement the modelling of the asset dynamics with a model of shocks persistence proposed by Arellano, Blundell and Bonhomme (2017). The results here show that income shocks at households trapped in dynamic structural poverty are less persistent compared with the income shocks of other households. As a consequence, households trapped in dynamic structural poverty facing positive income shocks, are temporary lifted out of monetary poverty, but the dynamics of structural income push them below the monetary poverty line again in the long run.

The third chapter makes visible the importance of income shocks when modelling income, and in particular poverty. For this reason, the fourth chapter focuses on measuring how well Colombian households are shielded against health shocks that may end up hurting the household capacity to accumulate productive assets and to generate income. Following this idea the fourth chapter of this thesis is called "The effect of different types of health insurance on health outcomes, medical care use, and risk protection: evidence from Colombia".

The health insurance system in Colombia is comprised by three types of insurance: individuals with private insurance, individuals on public funded health insurance for the poor and the uninsured population. This topic has been previously studied for Colombia, but in a different public policy setting.

Previous studies found that right after its introduction, the health insurance for the poor increased the use of preventive healthcare services when comparing them with individuals belonging to the private insurance, or those who were uninsured (Miller et. al, 2013; Camacho and Conover, 2013). This was due to the introduction of the public funded insurance for the poor in 1993 increasing access to healthcare services and risk protection for poorer households with little to no access for those services previously. For 2009, near-universal healthcare coverage was reached and with it, the goals of the public policy were adjusted accordingly. 2010 onwards the equalization of the scope and rights in healthcare services among the three health insurance types have been pursued. This was tested, using data for Colombia for the years 2010, 2013 and 2016 through a Fuzzy Regression Discontinuity design. The results show that in 2010 and onwards, there is no significant difference among the three possible health insurance types in outcomes related to health status, medical care use, risk protection against illness and behavioural distortions in urban areas. The result support the hypothesis of reaching the goal of the equalization in healthcare services, even for the uninsured population, who before 2009 were not entitled to most of healthcare services.

2.

References

[1] Arellano, Manuel, Richard Blundell and Stéphane Bonhomme. Earning and consumption dynamics: A nonlinear panel data framework. *Econometrica.* 85(3): 693-734. 2017

[2] Banerjee, Abhijit and Andrew Newman. Occupational Choice and the Process of Development. *Journal of Political Economy*, 101(2): 274–298. 1993

[3] Becker, Gary and Nigel Tomes. An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy.* 87: 1153–1189. 1979

[4] Carpio, Carlos, Michael Wohlgenant and Charles Safley. A Structural Econometric Model of Joint Consumption of Goods and Recreational Time: An Application to Pick-Your-Own Fruit. *American Journal of Agricultural Economics.* 90(3): 644-657. 2008

[5] Camacho, Adriana and Emily Conover. Effects of Subsidized Health Insurance on Newborn Health in a Developing Country. *Economic Development and Cultural Change.* 61(3): 633-658, 2013

[6] Carter, Michael and Christopher Barrett. The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies.* 42(2): 178-199. 2006

[7] Dasgupta, Partha and Debraj Ray. Inequality as a Determinant of Malnutrition and Unemployment: Theory. *The Economic Journal.* 96(384): 1011–1034. 1986

[8] Duncan, Greg, and Willard Rodgers. Has Children's Poverty Become More Persistent?. *American Sociological Review.* 56: 538-50. 1991

[9] Dutta, Swati. Identifying Single or Multiple Poverty Trap: An Application to Indian Household Panel Data. *Social Indicators Research.* 120(1): 157–179. 2015

[10] Foster, James. A Class of Chronic Poverty Measures, in *Poverty Dynamics: Interdisciplinary Perspectives*, ed. by Addison, Tony, David Hulme and Ravi Kanbur, 59–76. Oxford: Oxford University Press. 2009

[11] Foster, James, Joel Greer, and Erik Thorbecke. A Class of Decomposable Poverty Measures. *Econometrica*. 52(3): 761-66. 1984

[12] Galor, Oded and Omer Moav. From physical to human capital accumulation: Inequality and the process of development. *Review of Economic Studies*. 71: 1001–1026. 2004

[13] Galor, Oded and Joseph Zeira. Income Distribution and Macroeconomics. *Review of Economic Studies*. 60(1): 35–52. 1993

[14] Giesbert, Lena and Kati Schindler. Assets, Shocks, and Poverty Traps in Rural Mozambique. *World Development*. 40(8): 1594-1609. 2012

[15] Heckman, James and Stefano Mosso. The Economics of Human Development and Social Mobility. *Annual review of economics*. 6: 689–733. 2014

[16] Miller, Grant, Diana Pinto, and Marcos Vera-Hernández. Risk Protection, Service Use, and Health Outcomes under Colombia's Health Insurance Program for the Poor. *American Economic Journal: Applied Economics*. 5(4): 61-91, 2013

[17] Phillips, Peter and Donggyu Sul. Transition modeling and econometric convergence tests. *Econometrica*. 75(6): 1771-1855. 2007

[18] You, Jung. Risk, under-investment in agricultural assets and dynamic asset poverty in rural China. *China Economic Review*. 29: 27-45. 2014

Chapter 2. Welfare and poverty traps in urban Colombia: an approach from the asset dynamics perspective

Abstract

This paper examines the existence of poverty traps through household asset dynamics in urban Colombia. Using two different panel data surveys between 2007-2010 and 2010-2013 a variety of methods for estimating asset dynamic poverty traps were used. This is one of the few works that deal with the estimation of poverty traps in urban areas opening the door to a couple of technical problems, primarily related to the fact that in urban households the working members may be either employed or self-employed. The main result shows that even after considering this intrinsic characteristic of the urban households, there is no evidence in favour of the existence of a poverty trap in the 14 main Colombian cities, while the results are mixed for the other urban areas depending on the estimation method. This paper also finds that depending on the estimation method there may be one or three dynamic equilibria of the proxy for household assets. In the case of the existence of multiple equilibria the question on the mechanism that makes a household converge to a high or low equilibrium remains open.

Keywords: Poverty trap, Multiple equilibria, Single equilibrium, Asset dynamics, Welfare dynamics

JEL classification: I31, I32, J24, J44, J46, O11, O12, O17, C14, C18, C21, C23. C26

1. Introduction

The measurement of poverty is a relevant topic open to discussion, as it is a difficult task not only from a conceptual framework but also from the applied perspective. The usual poverty measurements have dealt extensively with the problem of static poverty comparing the household income with a given monetary poverty line (Foster, Greer, Thorbeck, 1984). Later this conceptual framework was extended to develop the idea of multidimensional poverty as the aggregation of static poverty indices across multiple dimensions beyond monetary, including but not limited to health, education and other dimensions considered relevant to individual welfare (Filmer and Pritchett, 2001; Bourguignon and Chakravarty, 2003; Maggio, 2004; Ravallion, 1994; Kai-Yuen, 2002). More recently dynamic poverty measurements have been introduce, mainly aggregating static poverty indices through time (Calvo and Dercon, 2007; Fuji, 2014 and Gradin, Del Rio and Canto, 2012).

Such poverty measurements are easy to implement as they require typically low quantities of information (usually limited to cross-sectional data) and are straightforward to construct. Nevertheless, this simplicity implies a lack of economic modelling on the poverty measurement. This is a determining characteristic, as it does not allow making precise policy evaluation, in the sense that besides the predesigned treatment-control environment is not possible to determine how public policies affect poverty. Even in the context of treatment-control experimental design, the conclusions on the effects of public policies are based merely on statistical models that do not allow understanding the individual's decision process, like the emergence of substitution/income effects once an individual becomes part of a public policy program. In this sense, Carter and Barrett (2006) propose a set of poverty measurements that allow not only to distinguish the structural foundations of poverty but also to give a glance at the long-term persistence of structural poverty.

The poverty trap models based in the macroeconomic context allow to classify households according to their long-term, persistent poverty status through a further understanding of the underlying patterns of asset dynamics and its relation with the income generating process of the household.

A wide range of applied papers have embarked in the search for poverty traps in different countries based on the Carter and Barrett (2006) theoretical framework. The idea of such literature is to find a proxy to the long run income of the household, i.e the income that is explained by the possession of productive assets, and then study the dynamics of such long run income. If a poverty trap exists in the Carter and Barrett (2006) fashion, an S-shaped structural income dynamics with multiple equilibria should be found. Not only that, but the lowest stable equilibria should be below the monetary poverty line. On the empirical side, the estimation of such poverty traps has been done mainly through two step methods, first estimating the long run income and second estimating the dynamic relationship among the long run income on two points of time (Carter and Barrett, 2006; Naschold, 2012 and Giesbert and Schindler, 2012) using non-parametric methods (Lybbert et. al , 2004; Barrett et. al, 2006; Quisumbing and Baulch, 2013) and semiparametric methods (Kwak and Smith, 2013; Gomez and Lopez, 2013; You, 2014).

Despite such extensive literature dealing with the topic, there are still many edges untouched. First, to the best of my knowledge the only paper trying to estimate such poverty traps in urban areas is Antman and McKenzie (2008). The authors develop a method to estimate dynamic panel data models in the presence of measurement error as they use a survey that only asks for labour income, while any poverty measurement by income should include all sources of household income. Second, estimating poverty traps for urban areas implies having to deal with some technical complications that rural areas do not face. On the one hand, in rural areas it is usually assumed that all household members work in the farm and therefore each household member faces the same income generating function. This is an unrealistic approach in urban areas as some household members may be self-employed while other may be employees. On the other hand, in rural areas all households are assumed to be self-employed working in their own farm. This again is an imprecise approach in urban contexts, as households may vary between situations where all members of the household are self-employed to situations where all of them are employees. This brings a problem first noted by Pissarides and Weber (1989) who found that self-employed households tend to underreport their income in tax forms and surveys in comparison with their employee counterparts.

In this context, the goal of this paper is to estimate the existence of poverty traps in Colombian urban areas following the theoretical framework of Carter and Barrett

(2006) and the two step estimation approach proposed by Barrett et. al (2006) but in this case taking into account the household composition between employed and self-employed members. This paper contributes to the poverty measurement literature in four main fronts: First it tests the existence of poverty traps in urban areas using the total household income and testing it with two different surveys for urban areas in Colombia. Second, it tests how the existence of poverty traps changes in respect to different ways of aggregating the assets across members of the household. Mainly the paper compares the results when using assets of the head of the household, average possession of assets among workers and total assets among workers. Third, this paper tests the existence of poverty traps when correcting for the income underreport of the self-employed households. Finally, the paper compares the conclusion on the existence of poverty traps among different estimation methods.

The results show that there is no a poverty trap in the Colombian cities. Despite finding evidence of highly nonlinear income dynamics with an S-shaped function as hypothesised by Carter and Barrett (2006) none of the stable equilibria is below the income poverty line, ruling out the existence of such poverty trap. The different methods of aggregation of assets and the correction for the income underreport of the self-employed households do not affect the general shape of the income dynamics, but do affect the number of statistically significant equilibria. Nevertheless, the estimated shape of the income dynamics changes according to the estimation method. The two-step procedure tends to find a single equilibrium while the more flexible estimation methods (non-parametric and semiparametric) tend to find two or more significant equilibria. This implies that in the 14 main Colombian cities, social mobility exists as households below the poverty line tend to move through time out of poverty. For the other urban areas (different from the 14 main cities) the evidence is more mixed, as the income dynamic function shows that in some cases there is low social mobility.

The remainder of this paper is organized as follows: section 2 presents the theoretical framework, section 3 makes a brief summary on the literature on poverty trap estimations using household data, section 4 describes briefly the two Colombian surveys used to estimate the existence of poverty traps in urban areas between 2007-2010 and 2010-2013; section 5 pin down the methods used in the paper. Section 6 shows the results of estimating the poverty trap presented in section 2 with the data described in section 4. Finally, section 7 presents briefly the conclusions and research agenda.

2. Theoretical Framework

On the one hand, the notion of poverty trap is directly derived from macroeconomic growth theory. As defined by Azariadis and Stachurski (2005) a poverty trap is a self-reinforcing mechanism that causes poverty to persist. According to Arunachalam and Shenoy (2016) the literature on poverty trap theories at a macroeconomic level can be classified according to the cause of such poverty trap in geographical (Krugman, 1991), imperfect credit (Matsuyama, 2004; Quah, 1996), and coordination failure (Murphy et al., 1989) causes. On the other hand, another set of theories focuses on households. Theories of occupational choice and lack of physical capital (Banerjee and Newman, 1993), human capital (Galor and Zeira, 1993), nutrition (Dasgupta and Ray, 1986), and contractual distortions resulting from moral hazard (Mookherjee and Ray, 2002) try to explain local inequality: why one family is poorer than another. Given that inequality within countries explains a large part of the global distribution of income (Bourguignon and Morrisson, 2002), the household poverty trap is no less important than the economy-wide poverty trap.

As Antman and McKenzie (2007) note such variety of poverty trap theories has lead to two different approaches when testing for poverty traps. One strand of the empirical literature has attempted to test particular theories of poverty traps. A second strand of recent literature has attempted to look directly at the dynamics of income, expenditure or assets to test for non-convexities. This paper follows this second approach, estimating the dynamics of the long run income. It follows the seminal work of Carter and Barrett (2006) which proposed an asset-based approach to distinguish a structural component of poverty, from poverty that may be overcome naturally with time, due to a systemic income growth process. Consequently this section reproduces various graphs, equations and arguments from Carter and Barrett (2006).

Such distinction on the persistence of poverty leads to different poverty measurements. According to Carter and Barrett (2006) taxonomy the poverty measurements can be divided into four classes: The first generation that relies in the comparison of household income (or expenditure) with a monetary poverty line. This method relies on cross sectional data. Such comparison allows to divide the population into poor and non-poor, while the repeated application of said measurement to cross sectional surveys through time would be a good description of the evolution of poverty.

The second generation of poverty measurements is based on panel data offering repeated observations of households or individuals over time. This allows to divide the population into three categories; those considered always poor, those considered transitory poor and those who are never poor (Carter and Barrett, 2006).

The third generation of poverty measurements acknowledges that the first and second generation are based only on monetary metrics -either income or expenditure- and ignores whether a household transitions in and out of poverty may be either structural or stochastic. Carter and Barrett (2006) propose to use the concept of asset based poverty line for being able to distinguish whether a household is poor due to structural conditions or stochastic shocks. The idea behind this approach is that the productive assets that a household possesses map into the income through certain income generating function, allowing to isolate which part of the income is due to productive assets -structural income- from the portion of the income which is only transitory. Those poverty measurements allow to isolate the structural poverty from the poverty due to stochastic events in a single period.

The fourth generation of poverty measurements should allow not only to distinguish the structural foundations of poverty but also to give a glance at the long-term persistence of structural poverty. In Carter and Barrett (2006) words, "the analysis based on the asset poverty line cannot [...] identify whether the currently structurally poor are likely to remain poor over the longer term, caught in a poverty trap, or whether some of the structurally non-poor may remain non-poor over the longer term". This kind of decomposition requires not only to be able to model the dynamics of income, but also to include an analysis of the dynamic evolution of the assets, which in the theoretical framework determine the evolution of the structural income (Carter and Barrett, 2006).

2.1 The Asset Based Approach

This section borrows from Carter and Barrett (2006). As defined previously, the asset poverty line allows to distinguish the households who do not have enough assets to generate a level of expenditure or income above the monetary poverty line. The relationship between assets and income (or expenditures) can be modelled using what Carter and Barrett (2006) call a "expected livelihood function". The difficulty in implementing this concepts comes from the necessity to estimate a livelihood mapping between assets and expenditures (or income) through some econometric model.

Following Carter and Barrett (2006) definition the dynamic asset poverty line distinguishes households caught in a long-term structural poverty trap from those expected to follow an upward trajectory, that is, the households that have structural economic mobility and therefore accumulate assets through time such that are able to leave poverty. This section explains the theoretical foundations for the dynamic asset poverty line as developed by Carter and Barrett (2006). Such asset poverty line is in simple words the "threshold at which accumulation dynamics bifurcate, leading to multiple dynamic welfare equilibria, including the possibility of a poverty trap" (Carter and Barrett, 2006).

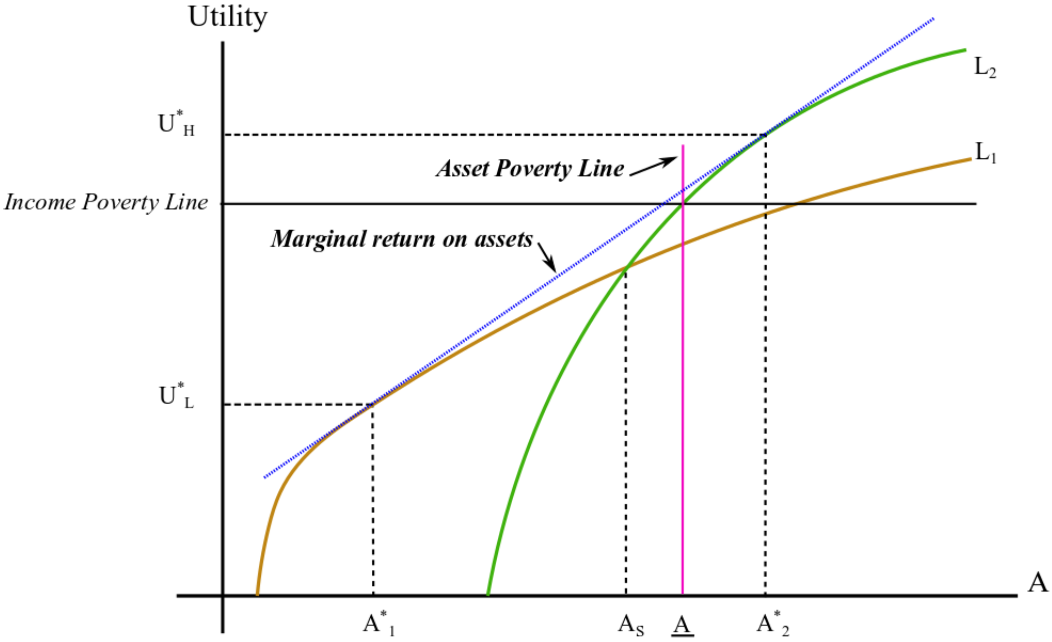
Similar to the country case, in the household case may also exist locally increasing returns to scale leading to multiple equilibria, and reducing the ability of poor households to catch up and converge at the same equilibrium than their richest counterparts. By definition, locally increasing returns implies a positive relation between wealth (measured in this context by the level of assets) and the marginal returns to assets. According to Carter and Barrett (2006) at the household level, such positive relationship between wealth and marginal returns can exist for at least three reasons. First, the income generating process may itself have increasing returns to scale. Second, high return production processes usually exhibits a fixed entry cost as a consequence of the need of a minimum project size such that only wealthier households can afford to adopt said high return production process. Third, lower wealth households may prefer to allocate their assets to reduce risk exposure, investing in productive activities with low expected returns and low risk, making marginal returns to wealth lower for lower wealth households.

Carter and Barrett (2006) illustrates the existence of the asset poverty line with an example. Let L_1 and L_2 be two productive activities in which a household may take part of. As usual, both production technologies are assumed to have decreasing returns to scale, but L_2 has a minimum scale of operation due to sunk cost. Figure 1 shows such pair of technologies and the optimal choice of household assets, which in this case coincides with the steady state value for the assets. In this example it is assumed that the household is restricted to only one of such technologies.

Let A_1^* be the asset steady state value for households confined to productive activity L_1, generating an income level of U_L^*. Let A_2^* be the equivalent for productive activity L_2, converging to the the steady state income U_H^*. Figure 1 shows this situation assuming that the asset poverty line, $\underline{\mathrm{A}}$, lies between A_1^* and A_2^* . If this is the case any household restricted to productive activity L_1 that consequently converges to equilibrium A_1^*, would be caught in a poverty trap despite the fact that a non-poor equilibrium exists in A_2^*. Given this theoretical situation what would be the optimal productive activity chose by the households?. If there are no other external constrains to the adoption of a given production technology, Figure 1 shows that, if on the one hand, a household has a level of assets between 0 and A_S, the optimal livelihood choice is productive activity L_1. On the other hand, if a household possesses a level of assets above A_S, L_2 would be the optimal livelihood choice. Despite both of these livelihood functions exhibiting diminishing returns, there are locally increasing returns in the neighbourhood of A_S, as the marginal return to assets just above A_S is higher than the marginal return to assets just below A_S (Carter and Barrett, 2006).

Carter and Barrett (2006) argue that this situations may be reflected in actual circumstances. For instance, in rural areas, households possessing more assets may adopt higher return crop varieties, or in urban areas, wealthier households may have access to better education resulting in white-collar employment rather than unskilled casual wage labour.

Figure 1: Asset poverty with multiple livelihood options



Source: Carter and Barrett (2006).

From a static point of view the poorer households might use productive technology L_1. Nevertheless, from a dynamic perspective the relevant question is if the locally increasing returns around A_S is an actual barrier on the asset accumulation process of the poorest households, impeding them to accumulate assets and cross over asset level A_S, which would allow them to catch up with the richest households. In other words whether or not household investing and accumulation behaviour would be induced by the low marginal returns generated under the technology L_1 when such household tries to accumulate assets beyond the optimal point A_1^*. In Carter and Barrett (2006) words: "A forward-looking household would know that while the marginal returns to further accumulation beyond A_1^* are low, increased accumulation has strategic value in moving the household closer to the asset level(s) where returns sharply increase". The household's best choice would be to borrow enough capital so that it could increase its asset possession to reach a higher return asset level.

In this context Carter and Barrett (2006) introduce the basic intuition of the dynamic asset poverty line. The authors argue that according to the distance of the households to the points A_1^* and A_S they may prefer to de-accumulate assets and settle down on the static optimal asset quantity A_1^* (if they are in fact closer to A_1^*), or accumulate assets up to A_S where increasing returns occur, if they are not far from A_S. Zimmerman and Carter (2003) identify such threshold as the Micawber frontier and define it as:

> ***A Micawber threshold, is the critical asset threshold below which it is no longer rational or feasible to pursue the autarchic accumulation strategy. If it exists, the Micawber threshold thus constitutes a dynamic asset poverty threshold, analogous to the static asset poverty line discussed before. Households whose assets place them above that threshold would be expected to escape poverty over time, while those below would not. One needs to identify this dynamic asset poverty threshold in order to disaggregate the structurally poor into those expected to escape poverty on their own over time through predictable asset accumulation and those expected to be trapped in poverty indefinitely.*** (Carter and Barrett, 2006 p.190)

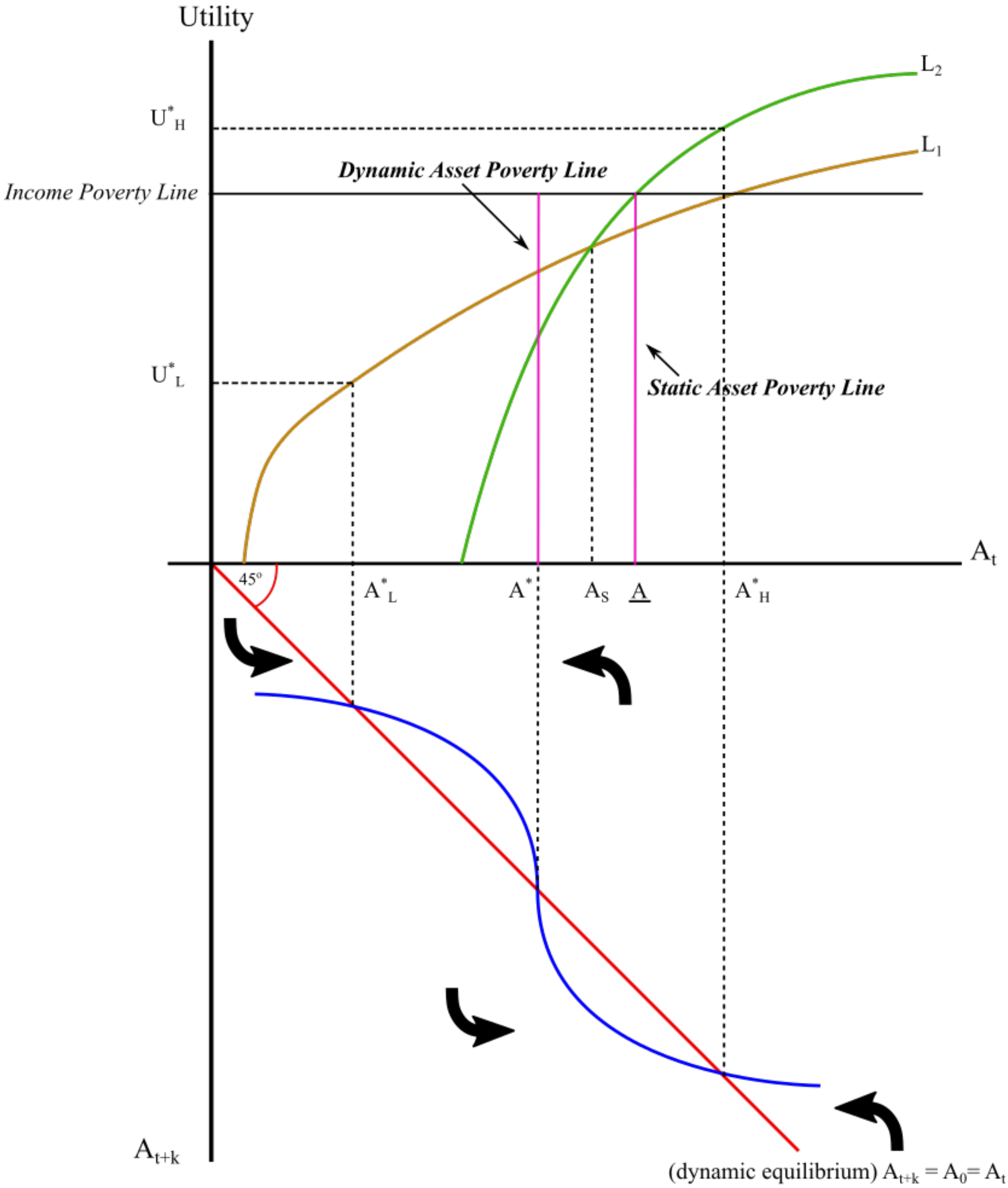
In this case the authors assume a given Micawber threshold (or asset dynamic poverty line) in the point A^*, and assume that $A^* < A_S$. The main implication of the idea previously described is that households with assets above A^* will accumulate assets, regardless of having lower marginal returns to asset than the returns obtained in the statical optimal, A_1^*. After reaching the point A_S (to be more precise when the household assets surpass A_S) the new statical optimal becomes to " switch to livelihood strategy L_2 and to grow to a steady state level of capital, A_2^* " (Carter and Barrett, 2006). On the contrary, households with assets between 0 and A^* would not find optimal to make the sacrifices needed to reach A_S, settling down in the steady state level of capital, A_1^*.

The top panel of figure 2 shows the situation previously described in the plane assets (A_t) on the horizontal axis and utility on the vertical axis. The bottom panel of the same figure shows what the implication of a hypothetical situation like this would be for asset dynamics, representing the households assets in t on the horizontal axis (A_t) and the household assets in a future period on the vertical axis (A_{t+k}). Carter and Barrett (2006) claim that in this example the difference between the income poverty line, the static assert poverty line and the dynamic asset poverty line can be seen. Indeed, the dynamic asset poverty line (A^*) do not coincide with the static poverty line ($\underline{\mathrm{A}}$), nor with the point where households optimally change in the static problem from the livelihood alternative L_1 to L_2 (A_S) and, as expected, is not related with the monetary poverty line.

As shown in the bottom panel of figure 2 and from a dynamic perspective, A^* is an unstable dynamic asset equilibrium. As with any unstable dynamic equilibrium, if a households possesses assets above A^* the dynamic evolution of the assets will make households to accumulate assets up to reaching the stable equilibrium A_H^*, yielding steady state utility U_H^* which is located above the income poverty line. On the contrary, if a household have initial assets between 0 and A^* will shed assets due to the dynamic of the asset accumulation process, and reach the asset steady state A_L^*, generating a utility level U_L^*, well below the income poverty line. In this case, given the stable nature of the equilibrium at A_L^*, this would act as an attractor, impeding the household to accumulate assets that allow them to produce a level of income above the poverty line. In conclusion, those households would be trapped in dynamic structural poverty.

The fourth generation of poverty measurements allows to distinguish people in transitory structural poverty from those trapped in dynamic structural poverty. This mechanism is described in Carter and Barrett (2006) words as “ in this particular case [...] ($A_L^* < A^* < \underline{\mathrm{A}}$), the structurally poor at any point time (those with assets below $\underline{\mathrm{A}}$) can be divided into those who will be persistently poor (A$<A^*$) and those who will eventually surpass $\underline{\mathrm{A}}$ on their way to the high level equilibrium, A_H^* (A^*<A<$\underline{\mathrm{A}}$)”.

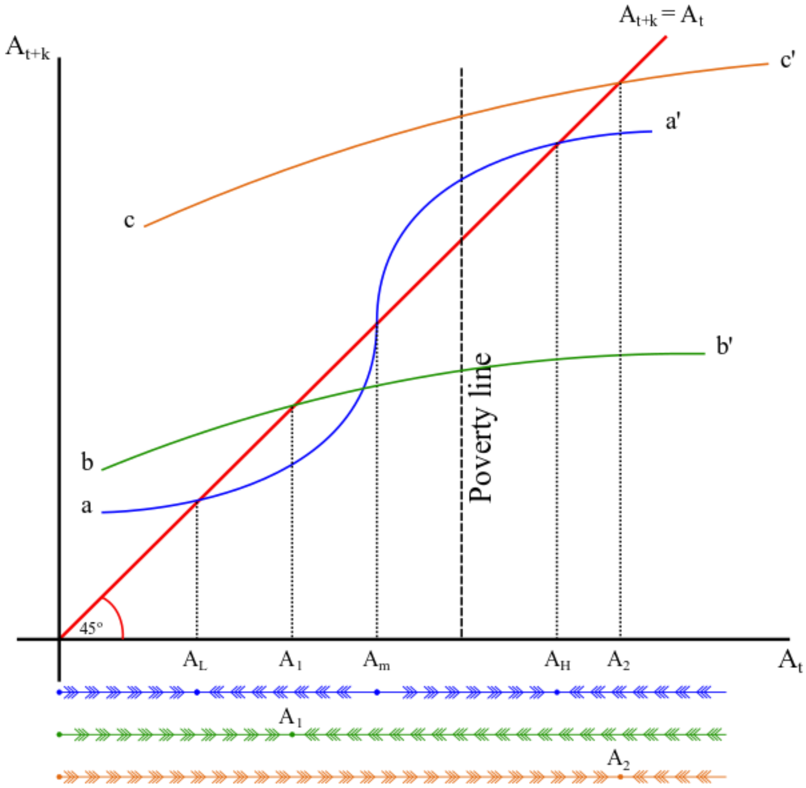
Figure 2: The dynamic asset poverty line



Source: Carter and Barrett (2006).

Carter and Barrett (2006) model consider only the possibility of a poverty trap with some multiple equilibria (two stable equilibria and one unstable) which would allow households to escape such trap under certain circumstances. Nevertheless, other authors recognise the possibility of the existence of such poverty trap even in the context of a single stable dynamic equilibrium. Dutta (2015) illustrates that situation with the Figure 3. It shows in the horizontal axis the household assets in a given initial period t, while the vertical axes show the household assets in a final period, $t+k$. In practice household asset dynamics can have any shape, and in particular may have only one stable equilibrium which may be either above poverty line, like in cc', converging to A_2, where no households are trapped in dynamic structural poverty. In contrast, if asset dynamics follow bb', the single steady-state is on A_1, below the poverty line, forcing all households following said dynamics into a dynamic structural poverty trap.

Figure 3: Different patterns of asset dynamics



Source: Dutta (2015).

In general, as Kwak and Smith (2013) point out there is no reason why, even in the case of the existence of multiple equilibria, the lowest equilibrium in figure 3 (A_L) must be below any commonly accepted poverty line. In the case of a single equilibrium it may be well below the poverty line (be in a poverty trap) or above the poverty line as shown by Dutta (2015). In the most extreme situations household asset dynamics may always be below the 45-degree line, in which case everyone is subject to a poverty trap converging to zero income (Antman and McKenzie, 2007).

An important feature usually overlooked in this literature is the analysis on the relationship between the asset dynamics and the 45-degree line. The idea here is directly extracted from macroeconomics: along the 45-degree line $A_t = A_{t+k}$, therefore if the estimated dynamics of assets is below this line, then $A_t > A_{t+k}$. This means that the household assets are shrinking and, assuming that the asset dynamics remain the same across time, the household assets would keep moving along the estimated dynamics (for instance aa') period by period, eventually reaching the line $A_t = A_{t+k}$. On the contrary, if $A_t < A_{t+k}$ the household assets would be somewhere above the 45-degree line continuously increasing and reaching the point where $A_t = A_{t+k}$ along the estimated dynamics.

The empirical literature on this topic usually operationalize A_t as the part of the household per-capita income (or expenditure) divided by the monetary poverty line in t, explained by assets in possession of the household. The mechanism previously describe is founded on the comparability of the variables in the two axis. In this sense, it would imply that the household per capita income divided by the poverty line in t (explained by assets) is comparable with the household per capita income divided by the poverty line in $t + k$ (explained by assets). This requires that the quotient of the per capita income to the poverty line represent the same in t and $t + k$; therefore, it is necessary to assume that the poverty line evolves through time at the same pace than the household per capita income. In fact, the construction process of the monetary poverty line includes this characteristic in its estimation, as will be explained later.

If this assumption does not hold, the line that represents the dynamic equilibrium in figure 3 may have any other angle, different than 45 degrees.

2.2 Poverty traps in urban areas

The first step in estimating a dynamic poverty trap is estimating some proxy for livelihood conditions. Some authors call it an asset index, while some others call it simply the long run income or structural income in the sense that it is the part of the total income that can be explained by the possession of productive assets.

Most of the literature estimate such livelihood condition proxy running a regression between the per capita income (or expenditure) of the household divided by the poverty line as left hand side variable, and the assets used to generate such income or expenditure as exogenous variables. The predicted value of the per capita income in such regression is interpreted as the "long run" or "structural" income. This type of approach is usually used to estimate such poverty traps for rural areas with only very few studies focusing on urban areas.

The empirical estimation of poverty traps based on those methods have a strong theoretical support for the rural areas. In the first place, because households who live in rural areas obtain most of their income from the agricultural production, making the relation between household income and physical productive inputs very strong. From a theoretical perspective estimating a function like this would just be estimating a profit function, considering each rural household as a different production unit. In second place, in developing and poor countries where the income inequality between urban and rural areas is higher, rural households usually have low opportunities for social intergenerational mobility, which in turn implies that most of the household members work as farmers in the production unit owned by the household. Then, when estimating the profit function for the farm it is realistic to assume that the household income is generated by a unique production function which combines all the members of the household as labour force with all the physical inputs as capital stock.

Nevertheless, this very same method of estimation rises some concerns when dealing with poverty traps in urban areas. First, as Antman and Mackenzie (2007) note "if the set of productive physical assets available for measurement is small (as in many urban contexts) or income shocks are persistent, the dynamics of the component of income which is predictable from assets may differ greatly from the dynamics of true income". These asset-based approaches appear do work well in rural settings where one well measured major asset, such as cattle, is the main source of income. However, such approach has yet to be tested in urban contexts. Second, the aggregation within households is itself a problem. The income (dependent variable) is easy to aggregate, as it is possible just to add up the value of all the income across members of the household and across activities using the same periodicity. The main problem is on how to aggregate the inputs, for instance, which education level to use as proxy to the human capital of the household: average education of working members, total education of working members or education of the head of the household. This in the case of every single member of the household having the same activity and therefore being able to assume that the income generating function of all members of the household is the same. If the members of the household engage in different activities and therefore different income generating functions (employee, self-employee, entrepreneur) the problem becomes worst. This rises a third concern. As the income generating function changes per activity, ideally, the income of each individual should be modelled jointly with the occupational choice.

This paper focuses on dealing with the aggregation problem when estimating poverty traps in urban areas. The reason for this is that household surveys in urban areas have little information on productive assets. Furthermore, having few information on productive assets is a problem only when most of the individual income is derived from capital-intensive entrepreneurial or self-employed activities. However most of the urban inhabitants derive their income from being an employee. In practical terms, this implies that the only capital stock they bring to the labour market is their own human capital in the form of education and work experience. From the perspective of the firm who hire such kind of workers, its own profit function would include as inputs the capital stock (owned and managed by the firm), the labour force and some measurement of human capital. From the perspective of the employee and the household, the main inputs they bring to the labour market are education and work experience, while the wage is the return for such input.

With respect to the simultaneity of income and occupational choice, this is a complicated issue in itself and it goes well beyond the reach of this paper.

There are few papers that have dealt in the past with the problem of aggregating human capital for the household and estimating the human capital return for the household as a whole. Yang and Yuying (2001) develop a two-sector model for measuring the returns of human capital when the household members engage in different activities. In this case the authors want to model the difference in returns to human capital between agricultural and non-agricultural sectors. This is a very similar case to the one I am concerned where households would have members typically working as self-employee or entrepreneur (similar to the agricultural sector) and employees (similar to other non-agricultural activities). The model has two types of factors. On the one hand, a quasi-fixed factor that links the two production activities. On the other hand, a factor which is activity-specific. The key characteristic of the model is that the choice of the quasi-fixed factor should satisfy a cross-activity, household resource constraint.

The authors model three profit functions. An aggregate profit function for the household across members and activities, and two other profit functions for each specific activity. They specify formal schooling and experience as the key components of family human capital that enhance efficiency. Education is measured as the total years of schooling of all household workers, and experience as the sum of workers' years of experience, which is approximated by age minus schooling minus 7. A justification for these stock measures is that, because education and experience are costly investments, the attainments of all household workers are expected to have positive returns. These specifications enable them to compute the total returns to human capital, from estimating the aggregate profit function.

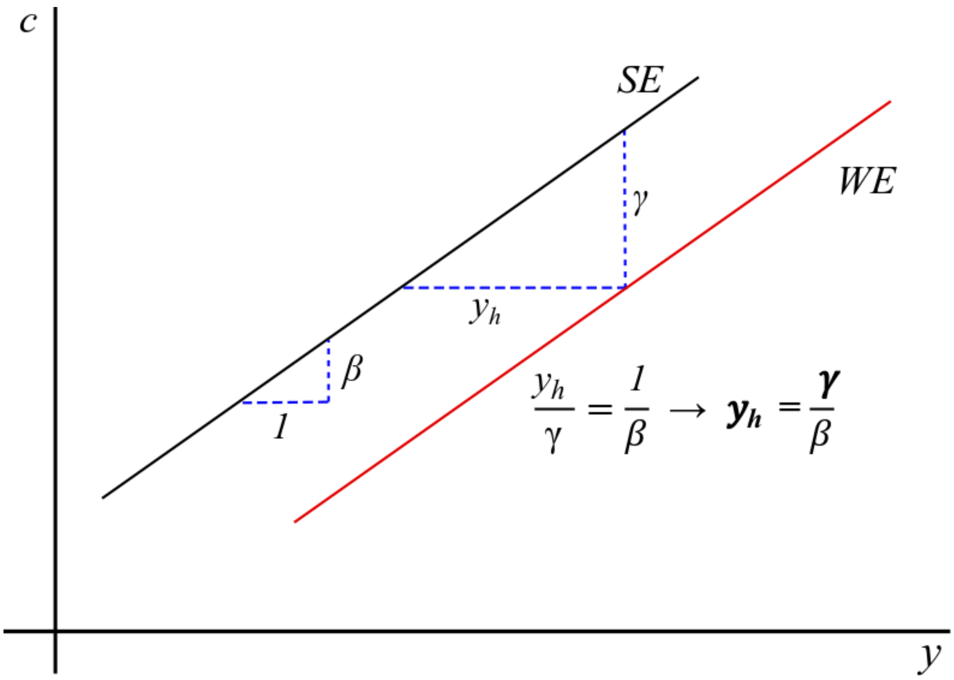
Laszlo (2005) estimates the returns to education for rural households in Peru who obtain a share of their income from non-farm self-employment endeavours. The main question of this paper is "how does the household stock of human capital affect household earnings when the household engages in more than one activity". For answering this, the article explores the aggregation of individual education levels within the household and how the aggregate household education affects total household income. In their model the earnings return to education depends on the way in which household schooling aggregates over individual schooling levels. Under this setting the solution of the model chooses aggregate household schooling in order to maximise household utility.

The author identifies two main effects. First, the worker effect representing the effect of schooling on wage through marginal productivity of labour. Second, what he calls an allocative effect which is related with the allocation efficiency of inputs within household production function, in this case, choosing optimally the allocation of human capital across activities. The estimation of such a model is done using as dependent variable the per capita household earnings from all sources of labour income and as exogenous variable the average years of education for household members aged 15 and above.

Another difficulty in the estimation process of poverty traps in urban context is related to the self-employed status of the household and was first pointed out by Pissarides and Weber (1989). In their seminal paper, they found that households where most of the members of the household are self-employees tend to underreport total income for purposes of tax evasion. Not only that, but a growing body of literature including Hurst, Li and Pugsley (2014), Kim, Gibson and Chung (2015), Engstrom and Hagen (2017) and Johansson (2005) found that this pattern of self-employed households underreporting their income is common, not only in tax forms, but also in household surveys.

Pissarides and Weber (1989) develop a method for finding in which proportion the self-employed households underreport their income. Following Engstrom and Hagen (2017), the Pissarides and Weber (1989) (PW now on) approach is shown in figure 4. Lets define $c = log(C^F)$, in words, the natural logarithm of total household consumption, while y denotes the natural logarithm of the disposable income, $y = log(Y^F)$. Assuming a linear relationship between the log of the consumption and the log of the disposable income, the figure shows the log-linear Engel curve for self-employed households (SE) and for wage earners (WE). The difference between the point where the curve for the wage earners crosses the vertical axis and the point where the curve for the self-employed crosses the vertical axis is an estimation of the amount of income underreporting of the self-employed in comparison with the wage earners.

Figure 4: Engle curves for for wage earners (WE) and self-employed (SE)



Source: Engstrom and Hagen (2017).

Engstrom and Hagen (2017) describe four assumptions needed for such a model to be reliable. First, the elasticity of consumption with respect to income, β, is equal for the two groups. This is illustrated by the curves having the same slope. Second, there is no systematic misreporting of expenditures between the two groups. The item of expenditure that most likely fulfils this assumption is food. There is little reason to lie about food consumption and it is also easy to report. Third, on the one hand, self-employed households underreport their income by a constant factor. On the other hand, wage earners report their income truthfully. Fourth, individuals who misreport their income in surveys do it in the same way to the tax authorities, as this was the situation originally studied by PW.

Using data for self-employed households and wage earners, Engstrom and Hagen (2017) propose to estimate the degree of underreporting among the self-employed through this equation:

$$c_{it} = X_i\alpha + \beta y_{it} + \gamma SE_{it} + \epsilon_{it} \tag{1}$$

here i, t represent the household and year identifiers respectively. X_i is a set of variables determining consumption, while SE_{it} is a variable taking on the value of one if the household is self-employed and zero otherwise. ϵ_{it} is an error term assumed to be randomly distributed. The parameter of interest is γ, which after some transformations captures the degree of income underreporting by th self-employed households and in this particular setting captures the difference in intercept between the two Engel curves. Let κ be the fraction of true income reported by the self-employed. After some trivial algebra, it can be shown that $\hat{\kappa} = \exp(-\hat{\gamma}/\hat{\beta})$, therefore, $\hat{\kappa}$ can be estimated from equation 1. Form here, 1-κ would be the proportion in which the self-employed households underreport their income.

Pissarides and Weber (1989) work is based on the idea that the income measure that should be included in this model is some proxy for the permanent income instead of current income. Nevertheless Engstrom and Hagen (2017) argue that measures of permanent income are scarce resulting in the use of current income which in turn, as showed by them, overestimate the degree of income underreporting by self-employed households.

Hurst, Li and Pugsley (2014), Kim, Gibson and Chung (2015) and Johansson (2005) have used instrumental variable estimations to use as proxy for the permanent income. Nevertheless, Engstrom and Hagen (2017) argue that in general "finding instruments that affect consumption only through permanent income is difficult, plus the impossibility to directly test the exclusion restriction without access to permanent income".

3. Empirical Literature Review

Chart 1 shows in an schematic way an overview of the papers that have been written on the topic of the estimation of poverty traps using household data. The chart shows the country in which the study was conducted, the estimation and econometric specification and the main result in terms of whether the authors found the existence of a poverty trap. Some of this papers will be used as further reference and explored deeper in the methodological section (section 5).

Summing up, the results of these studies are mixed in terms of the existence of a poverty trap itself and the number of equilibria found. In particular, on the one hand, for the rural areas of India, Ethiopia, México, Mozambique, Bangladesh and El Salvador a single equilibrium conducting to a poverty trap was found. On the other hand, multiple equilibria with one of the equilibrium leading to a poverty trap were found for rural areas on China, South Africa, Madagascar and Ethiopia. Finally, only one study found no poverty trap in rural China. When studying the literature carefully it can be concluded that the results are not robust to the estimation method (parametric, semi-parametric and non-parametric), the specification for the estimation of the household assets (linear or non-linear) and the measurement of household productive assets.

Chart 1: Literature on the estimation of poverty traps using household data (1)

Author (Year)	Area	Data and Span	Estimation Method	Results
Arunachalam and Shenoy (2017)	Rural	India. Additional Rural Incomes Survey, Rural Economic Development. 1969, 1982, 1999.	Developed its own poverty trap detecting method based on the idea that if far from a low attractor the probability of a negative income shock should decrease in a two equlibria environment. Non-parametric regression.	Single equilibrium not in a poverty trap.
Zhou and Turvey (2015)	Rural	China.Long household data panel. 1989-2009	4th degree polynomial in a 3SLS correcting for endogeneity of income with education	Not evidence of poverty trap
Dutta (2015)	Rural	India. India Human development Survey. 1993,2005	Local polynomial regression with Epanechnikov kernel weights and linear mixed models for estimating multidimensional poverty traps (Iliteracy and undernutrition).	Across all the states there is a single dynamic asset equilibrium. The most deprived states are in a poverty trap.
Mukasa (2015)	All	Uganda. Uganda National Panel Survey. 2005,2009-2011	Parametric methods (System-GMM with cubic polynomial), non-parametric methods (LOWESS and local polynomial regression), and semi-parametric methods (Ruppert splines)	Not evidence of poverty trap, but conditional convergence depending on the initial conditions of the household.
You (2014)	Rural	China. China Health and Nutrition Surveys. 1989-2006.	Two step procedure following Adato, Carter and May (2006) and semi-parametric methods.	Households risk management, together with multiple equilibria in agricultural asset dynamics can trigger dynamic asset poverty.
Kwak and Smith (2013)	Rural	Ethiopia. Ethiopian Rural Household Survey. 1994, 1999, 2004.	Parametric GMM-FE, Local linear Regression, Partial linear model with RE, Bayesian penalised spline. Quantile regressions	Single equilibrium from 1994 to 1999. Such equilibrium moved upwards on the period 1999 to 2004.
Gomez and Lopez (2013)	Rural	México. Mexican National Rural Household Survey. 2002-2007.	Multi-step approach following Adato, Carter and May (2006). Semi-parametric model.	Single low equilibrium with a low, or even null, asset mobility.
Maxwell et al. (2013)	Rural	Ethiopia. Livelihoods Change Over Time Survey. 2011-2013.	Locally weighted (non-parametric) regression.	Single equilibrium towards a poverty trap.

Chart 1: Literature on the estimation of poverty traps using household data (2)

Author (Year)	Area	Data and Span	Estimation Method	Results
Quisumbing and Baulch (2012)	Rural	Bangladesh. Various IFPRI Surveys. 1994-2007.	RE panel with second order polynomial, OLS, Locally Weighted Regression	Low-level equilibrium or poverty trap. Limited evidence for the existence of multiple equilibria.
Carter and Lybbert (2012)	Rural	Burkina Faso. International Crop Research Institute for the Semi-Arid Tropics. 1981-1985.	FE, non-parametric regressions by quantile and Hansen threshold estimation method.	Test the behavioural implications of the existence of a poverty trap, instead of test for the existence of a poverty trap directly. Finds evidence of multiple asset smoothing regimes as predicted by poverty trap theory.
Zerfu (2012)	Rural	Ethiopia. Ethiopian Rural Household survey. 1994-2004.	Arellano-Bond with cubic term and Local Polynomial Regression.	Single equilibrium poverty trap while controlling for the institutional framework.
Giesbert and Schiendler (2012)	Rural	Mozambique. Trabalho de Inquérito Agrícola. 2002-2005.	Multi-step approach following Adato, Carter and May (2006) controlling for exogenous variables.	Poverty trap with specific group convergence below the poverty line.
Naschold (2012)	Rural	India. International Crop Research Institute for the Semi-arid Tropics' Village Level Studies. 1975-2003.	Multi-step approach on the line of Adato, Carter and May (2006) using penalized spline.	Single equilibrium with a poverty trap.
Hien (2011)	All	Vietnam. Vietnam Household Living Standards Survey. 2004-2008.	Hansen's threshold estimator and quantile regression.	Multiple equilibrium. The evidence of an upper equilibrium is more significant than a lower equilibrium.
Kwak and Smith (2011)	Rural	Ethiopia. Ethiopian Rural Household Survey. 1994, 2004.	Bayesian penalized spline looking for poverty traps on literacy and nutrition.	Poverty trap in each single dimension, but not multidimensional poverty trap.

Chart 1: Literature on the estimation of poverty traps using household data (3)

Author (Year)	Area	Data and Span	Estimation Method	Results
Antman and McKenzie (2007)	Urban	Mexico. National Survey of Urban Employment. 1987, 2007.	Panel Data model wih measurement error. Locally weighted regression.	Non-linearities in labour income but not a poverty trap.
Adato, Carter and May (2006)	Rural	South Africa. KwaZulu-Natal Income Dynamics Study. 1993, 1998.	Multi-step approach. First finds a proxy for a wealth asset index estimating a random effects model using as dependent variable the household per-capita income. The predicted dependent variable is a proxy for structural income. The second step is estimating a non-parametric model between the structural income and its own lag for measuring whether the dynamics of the household' assets shows non-linearities.	Poverty trap evidence (two stable equilibria).
BarretT et al. (2006)	Rural	Kenya and Madagascar. Rural Markets, Natural Capital and Dynamic Poverty Traps in East Africa. 1989-2002.	Non-parametric LOESS model. Parametric model on a fourth order polynomial.	Non-linearities in income dynamics with a poverty trap.
Lybbert et al. (2004)	Rural	Ethiopia. Desta's data. 1980-1997.	Non parametric model: Nadaraya-Watson estimates using Epanechnikov kernel.	S-shaped income dynamics with a poverty trap.
Rodriguez and Gonzales (2004)	Rural	El Salvador. Data from micro finance study in Ohio-State University. 1995-2001.	Arellano-Bond using one lag and up to a fourth degree.	Single dynamic equilibrium with a poverty trap.
Jalan and Ravallion (2002)	Rural	China. 1985-1990.	GMM.	Evidence of a poverty trap.

4. Data

This study estimates the existence of poverty traps in urban areas in Colombia. For such task a panel data following urban households for more than one period is needed. There are only two panel data surveys that follows urban households in Colombia: First, the Fedesarrollo Longitudinal Social Survey (FLSS) and second, the Colombian Longitudinal Survey by Universidad de los Andes (CLSA). This section shows some basic descriptives of each survey and discuss their representativeness.

4.1 Fedesarrollo Longitudinal Social Survey

This study uses a subset of the Fedesarrollo Longitudinal Social Survey (FLSS) that followed urban households in Colombia from 2004 to 2010. The first round of the Fedesarrollo Longitudinal Social Survey (FLSS) was conducted in 1999, and from 2004 to 2010 it followed a group of households for 6 consecutive years (Fedesarrollo, 2010). This survey allows the characterization of urban households through time in aspects as dwelling quality, welfare conditions, demographic conditions, health, education and labour market. Since 2007 it also asks for a wide range of shocks suffered by the household during last year including health shocks, labour shocks, crime victimization, and different ways of coping with those difficulties like buying insurance, savings or access to credit. In 2008 and 2009, additionally to those questions it asks for the probability of suffering a shock in the next year and -in case of suffering- the ways in that the household has planned to deal with it.

In 2007 the survey sampled the urban population of Bogotá, Bucaramanga and Cali. Thereafter the universe was expanded also to the cities of Medellín, Barranquilla, Manizales, Pasto, Pereira, Cúcuta, Ibagué, Montería, Cartagena y Villavicencio. In 2010 -the last year when the survey was conducted- it was representative for the urban population for those 13 cities, which accounts approximately for 39.2% of the Colombian population for that year.

As almost every probabilistic panel survey the FLSS is exposed to attrition due to failing to follow the same set of households during the whole period, so that the sampling was designed as a rotating panel where a certain part of the original sample was replaced with new observations every year. According to the technical documentation in 2010 just 57.9% of the original sample in 2008 was still observed. In the same way, only 43.2% (802 households) of the original sample in 2007 is observed in 2010. This is an important characteristic to consider, as some of the econometric models require having observations for certain variables across all the periods of time (a balanced sample), implying that the survey has 802 households that constitutes a balanced sample. Nevertheless it is not expected that all the 802 households have information for all the variables necessary to implement the econometric models described, so the final balanced sample of households that have information on the relevant variables for the years 2007, 2008, 2009 and 2010 is 684.

There is a remarkable tradeoff between using the panel data sample from the years 2007-2010 or 2008-2010. In the first case, the sample will follow only the cities of Bogotá, Bucaramanga and Cali and 802 households, while in the second case it would use 13 cities and 2609 households.

As one of the main goals of this paper is to estimate changes in long run income, we used the panel data with the longest time dimension, but the smallest cross section sample. Therefore, the final sample covers the years 2007-2010 and the cities of Bogotá, Bucaramanga and Cali. It is representative of the population of these three cities and the 22.3% of the total population in Colombia.

A key aspect of this model when implemented empirically is the monetary poverty line. In this case I am not concerned with comparing period by period the monetary poverty line with the household income directly; nevertheless it should be used as a reference when estimating a model like the one proposed in figure 3. In Colombia the monetary poverty line is estimated by the National Institute of Statistics (DANE in spanish) using expenditure surveys. With this surveys the value of a "basic food basket" is estimated, and said value corresponds to the extreme poverty line (MESEP, 2012). To obtain the monetary poverty line, an estimation of the proportion of food expenditure in total expenditure is needed which can also be performed with the expenditure survey. Once this proportion is estimated (called Orshansky coefficient), it is multiplied by the value of the extreme poverty line and the result is the monetary poverty line (MESEP, 2012). For this period, the monetary poverty line was estimated with an expenditure survey for the year 2007 and updated year by year by the National Institute of Statistics using the consumer price index for low income households. I use the official monetary poverty line values year by year in this work.

Row 1 of figure 5 shows the kernel density estimates of the adult equivalent per capita income per household divided by the poverty line[1] in the years 2007 and 2010 for the FLSS. Both graphs show the poverty line as dotted and both show a skewness to the right as it is expected from any income distribution, where values of the income tend to be concentrated in lower quantiles while the further the income is from zero, the less values tend to be after the highest peak. In the case of the existence of a poverty trap, or in general, the existence of two equilibria in the income generating process it would be expected to see a bimodal density, which is the case in the 2007 adult equivalent per capita income divided by the poverty line but not for the same variable in 2010. In both cases the mode of the distribution is above the poverty line.

The left graph on the third row of the figure 5 shows the scatter plot between the adult equivalent per capita income per household divided by the poverty line for both years. Despise the existence of a general positive relation among the two variables, hypothesizing that such relation is linear is, at least, reckless. As Carter and Barrett (2006) point out, on one hand, estimating a parametric model for such kind of relation is not flexible enough for capturing the non-linearities that should be allowed. On the other hand, a full non-parametric model would allow a more flexible functional form but would not be able to distinguish whether what exists is a nonlinear relationship instead of just heterogeneity in the income due to non-observable characteristics.

Table 1A show the basic descriptives of the variables used in the econometric models for the FLSS. First, it is important to note that I am only considering households that didn't move among cities between the two rounds of the survey, so the distribution of the households in the whole period when the survey was conducted remains

[1] The per capita income adjusted by adult equivalence and standardized by the monetary poverty line (Y_{it}) is calculated based on the adult equivalence scale estimated by Muñoz (2014) for Colombia:

$$Y_{it} = \left(\frac{\text{Total Household Income}_{it}}{1+0.7089*(\text{Adults}_{it}-1)+0.6822*\text{Children}_{it}+0.6628*\text{Teenagers}_{it}}\right)\left(\frac{1}{\text{Poverty line}_t}\right)$$

where Total Household Income$_{it}$ is the household income from all sources and all members, for the household i at period t. Adults$_{it}$ is the number of household members 18+ years old for the household i at period t, Children$_{it}$ is the number of household members between 0 and 7 years old and Teenagers$_{it}$ is the number of household members between 8 and 17 years old. Poverty line$_{it}$ is the urban poverty line for the year t.

the same. 50.3% is located in Bogotá, 45.3% is located in Cali and the remaining 4.4% is living in Bucaramanga. The employment rate among household members decreases in 2 percentual points, from 49.8% to 47.9%. The age of the head of the household increases by approximately 3 years as is expected, while the education of the head of the household increases only in 0.3 years of education. This little increment in head of household's education is due, mainly, to the average age of the head of the household. The proportion of members between 0 and 12 years old decreases from 15.8% to 11.7%, which is an expected characteristic in urban populations with low birth rates, and, which is complemented with an increase in the proportion of members of 62+ years old which goes from 17.5% to 21.8%. The per capita income adjusted by adult equivalency and standardized by poverty line increases from 1.9 to 2.5 between 2007 and 2010. In other words, the average income of the household changes from 1.9 times the poverty line to 2.5 times the poverty line.

With respect to the exogenous variables that are aggregated over different household members, we have that the total age of working members decreases from 61.5 to 58.5 years old on average, which would be consistent with younger people coming into the labour market while the elders getting retired. Consistently with this finding, the total education of the workers decreases also from 19.18 to 18.73 years. When dealing with the averages over the worker members we found that the average age and average education increases. This result is contradictory with the totals due to not all the households having income from work, but some of them have it from members who are 62+ years old. If this is the case, when taking averages over education and age, the denominator of such calculation is zero, implying that those households become a missing value. Finally, the number of self-employed members decreases from almost 1.1 to 0.9.

4.2 Colombian Longitudinal Survey by Universidad de los Andes

The Colombian Longitudinal Survey by Universidad de los Andes (CLSA) had its first round in 2010 and its second round in 2013. On its first round, it interviewed in total 10.164 households with representativeness on five regions (Atlántica, Pacífica, Central, Oriental and Bogotá). Of those 10.164 only 9.830 households were intended to follow. Out of those 9.830 households (5.275) 53.66% are located in urban areas, while the remaining 4.555 (46.3%) are located in rural areas (Universidad de los Andes, 2014). In the second round, only 8.848 households could be interviewed which leads to an attrition rate of 9.9%. Of those 8.848 households, 4430 are urban, making this the final number of households in urban areas in the panel. Unfortunately, not all those households have the required variables for the econometric models I am running. This leave me with a final effective sample of 2.718 households, of which 51.8% belong to the 14 main cities (Bogotá, Bucaramanga, Cali, Medellín, Barranquilla, Manizales, Pasto, Pereira, Cúcuta, Ibagué, Montería, Cartagena, Santa Marta and Villavicencio) and 48.1% belong to other urban areas.

In terms of the structure of the survey it is very similar to the FLSS as it asks for dwelling quality, welfare conditions, demographic conditions, health, education, labour market, shocks and mechanisms for coping with those shocks. Nevertheless, different from the FLSS, the CLSA only asks for education and labour market variables for the head of the household and its partner in 2010, ruling out the possibility of using different aggregations over the variables of the workers of the household when estimating the econometric models.

Row 2 of figure 5 shows the density estimations for the adult equivalent per capita income per household divided by the poverty line in the years 2010 and 2013 for the CLSA. In this case, none of the two densities show a bimodal pattern, and, while in 2010 the poverty line seems to be very close to the mode of the estimated density, in 2013 the poverty line is below the mode of the density. In both cases the estimated density has the expected pattern when dealing with per-capita income. As with the FLSS, nothing can be said from the scatter plot between the income variables in 2013 and 2010, except that there exist a positive relation among them, that seems to be either nonlinear or very heteroskedastic.

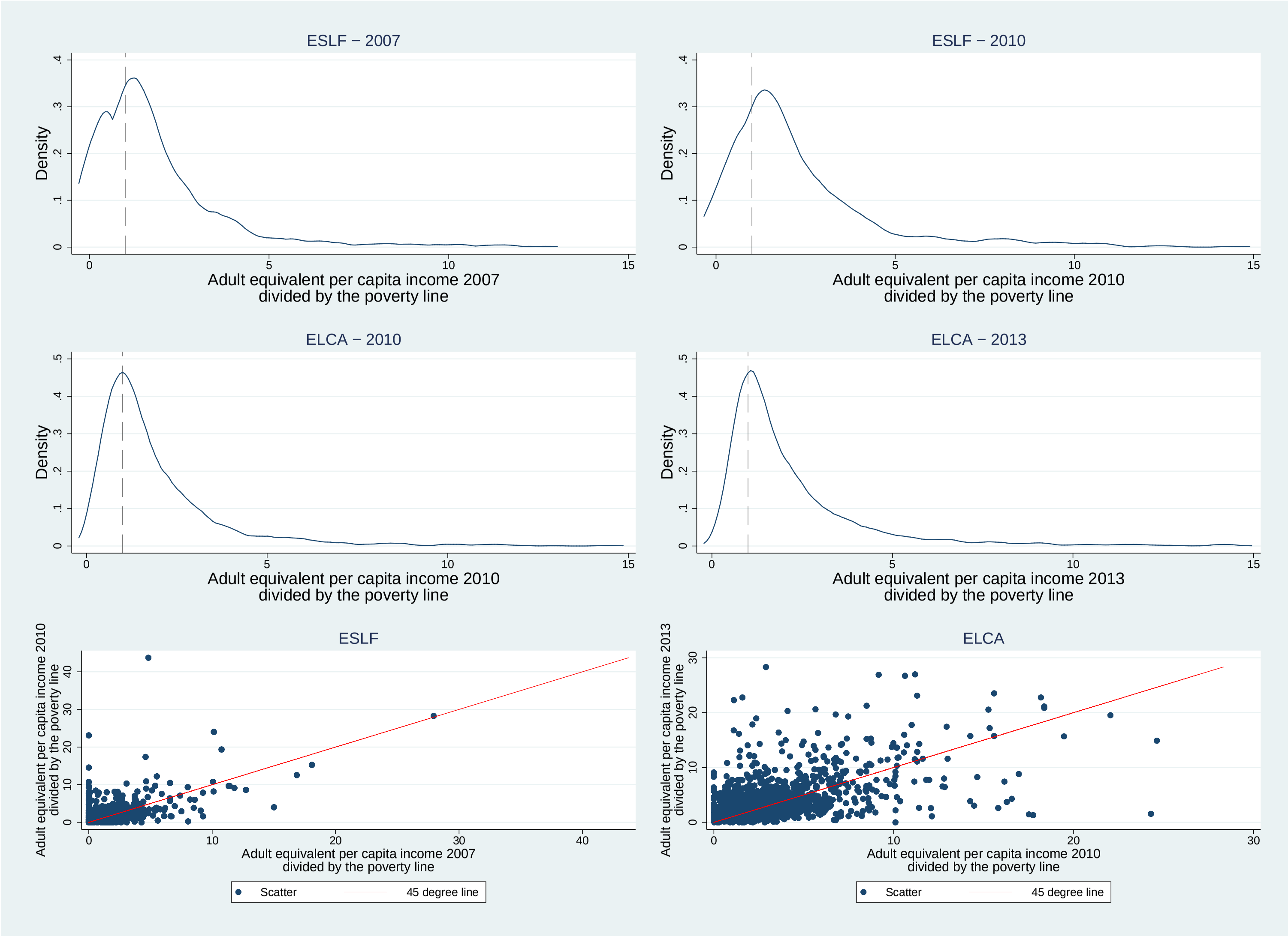
Table 1B shows the descriptives of the variables used for estimating the econometric models. As expected, the age of the head of the household increased in around 3.2 years, from 42 years old in 2010 to 45.22 years old in 2013. The proportion of members between 0 and 12 years old also decreases from .23 to 0.20 which is predictable as we are dealing with urban households where low birth rates are expected. Therefore, on average, is expected that after three years, more people become older than 12 years old, than the new-borns in the household. This is consistent with the proportion of members 62+ years old increasing from 0.05 to 0.06. The adult equivalent per capita income per household divided by the poverty line increases from 1.95 to 2.32 while the education of the head of the household increases from 11.21 to 12.29. The proportion of self-employed heads of the household increases a little from 52.1% to 52.8%.

As the CLSA only asks for labour variables for the head of the household and its partner, such variables are also included in the econometric models for the partner of the head of the household. All the changes in the variables of the partner of the head of the household show a similar behaviour to those of the head of the household. The education of the partner of the head of the household increases from 11.20 to 12.46, the age of the partner increases from 38.55 to 42.49 and the proportion of partners with a self-employed status decreases from 0.57 to 0.54.

As a final remark, it is important to say that the FLSS and CLSA are not comparable given that they are dealing with different periods of time and different samples. The FLSS uses only three main cities in Colombia while the CLSA includes 14 main cities and 73 other urban areas in Colombia, so it is expected that the sample behaves differently between the two surveys. A way to make the figures comparable would be to select the same three cities on the CLSA and then do the descriptives only for the same three cities that exists in the FLSS. Nevertheless, we are not able to do this given confidentiality agreements with the households taking part on the CLSA.

Figure 5

Density and scatter plots of the Adult Equivalent per Capita Income Divided by the Poverty Line FLSS (2007 and 2010) and CLSA (2010 and 2013)



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS) for the years 2007 and 2010 and the Colombian Longitudinal Survey by Universidad de los Andes (CLSA) for the years 2010 and 2013. Note: 1./ First and second row shows the kernel density estimates of the adult equivalent per capita income divided by the poverty line using epanechnikov kernel and data driven bandwidth selection method. The first row uses the FLSS data and the second row uses CLSA data. 2./ Third row shows a dispersion graph among the income in the initial year and the income in the final year with a 45 degrees line using the FLSS data (Left) and CLSA data (Right).

Table 1A: Fedesarrollo Longitudinal Social Survey
Descriptive statistics. 2007 and 2010
Household level

	(1)	(2)	(3)	(4)	(5)	(6)
	year 2007			year 2010		
VARIABLES	N	mean	sd	N	mean	sd
Bogota	684	0.503	0.500	684	0.503	0.500
Cali	684	0.453	0.498	684	0.453	0.498
Employment rate among household members	684	0.498	0.286	684	0.479	0.297
Age of head of the household	684	52.93	14.03	684	55.64	13.91
Education of head of the household	684	10.89	4.686	684	11.19	4.736
Proportion of members between 0 and 12 years old	684	0.158	0.184	684	0.117	0.166
Proportion of members 62+ years old	684	0.175	0.283	684	0.218	0.314
Percapita income adjusted by adult equivalency and standardized by poverty line	684	1.922	2.331	684	2.526	3.215
Age of working members (total)	684	61.54	42.43	684	58.54	42.60
Education of working members (total)	684	19.18	14.57	684	18.73	15.00
Age of working members (average)	587	40.77	10.97	553	41.95	11.30
Education of working members (average)	587	12.32	3.872	553	13.00	4.013
Number of self-employed household members	684	1.085	1.063	684	0.904	0.963

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS).

Table 1B: Colombian Longitudinal Survey by Universidad de los Andes
Descriptive statistics. 2010 and 2013
Household level

	(1)	(2)	(3)	(4)	(5)	(6)
	year 2010			year 2013		
VARIABLES	N	mean	sd	N	mean	sd
Age of head of the household	2,718	41.99	10.56	2,718	45.22	10.55
Proportion of members between 0 and 12 years old	2,718	0.233	0.204	2,718	0.208	0.196
Proportion of members 62+ years old	2,718	0.0511	0.132	2,718	0.0637	0.153
Percapita income adjusted by adult equivalency and standardized by poverty line	2,718	1.956	1.968	2,718	2.328	2.607
Education of head of the household	2,718	11.21	4.688	2,718	12.29	4.771
Education of head of household's partner	1,869	11.20	4.331	1,955	12.46	4.333
Age of head of household's partner	1,982	38.55	11.19	2,076	42.49	12.07
Self-employed head of the household	2,718	0.521	0.500	2,718	0.528	0.499
Self-employed head of household's partner	1,056	0.572	0.495	1,209	0.545	0.498

Source: Author's estimations using Colombian Longitudinal Survey by Universidad de los Andes (CLSA).

5. Methodology

From a theoretical stand point this paper follows the framework of Carter and Barrett (2006). I look for a poverty trap with three equilibria: two stable and one unstable. From the empirical side of the topic, the estimation of dynamic poverty traps has been done following a wide variety of econometric methodologies as shown in the empirical literature review section. Such methodologies range from simpler models like OLS, IV, any kind of data panel models -including FE and dynamic-, GMM, and ending up in non-parametric or semi-parametric regression models. In this section, I discuss the econometric methodologies used to estimate dynamic poverty traps in this paper.

As Barrett et.al (2006) point out, the disentangling of the causes of poverty implies being able to distinguish whether a household exits poverty due to structural causes, like capital accumulation, which would allow them to remain out of poverty permanently, from the cases where the household merely enjoys a temporary exit from poverty due to increases in the transitory income.

This conceptual framework has implications on the empirical strategy that should be used to estimate the existence of such poverty traps. In particular, Barrett et.al (2006) proposes to separate the structural or long term income from the transitory or short term income estimating the current observed income as function of the productive assets used to generate such income (estimate the so-called income generating function). After having this estimation, the predicted income may be considered the structural income, as it is the part of the observed income that can be explained only by productive assets.

This process can be repeated for different periods of time (different months, years, trimesters) regressing the current income on the current possession of productive assets for finding the structural income on each period. After having the structural income in any two points of time it is possible to estimate the non-linear relationship between the structural income in period $t + k$ and the structural income in period t. As Carter and Barrett (2006) argue, this last step should be made in a flexible way, such that the current structural income can take any functional form when explained by the lagged structural income. For this reason, most authors decide to estimate this model using non-parametric regression methods.

Here I have used a variety of methods for testing the robustness of the results to the estimation method. First, following Lybbert et. al (2004), Barrett et. al (2006), Quisumbing and Baulch (2012), Maxwell at. al (2013) and Mukasa (2015), I run a Kernel-weighted local polynomial smoothing regression of the adult equivalent income divided by the poverty line in 2010 (2013) using data of the FLSS (CLSA), on the adult equivalent income divided by the poverty line in 2007 (2010). This first model is merely descriptive, as it is only a broad approach to the nonlinear relationship among the current observed income in the two periods of time. No further conclusions can be drawn from this descriptive exercise as I am using the observed income, including its transitory component.

Second, following Barrett et.al (2006), Adato, Carter and May (2006), Naschold (2012), Giesbert and Schindler (2012), Gomez and Lopez (2013) and You (2014), I implemented a two steps procedure. In the first step, I estimated the long run income using OLS, explaining the structural income as a function of a set of variables that I consider relevant in the income generating function for urban areas. This was done separately for 2010 (2013) and 2007 (2010), allowing different coefficients for the explanatory variables in both years. In the second step, I run a Kernel-weighted local polynomial smoothing regression of the structural income (i.e predicted income) in 2010 (2013) explained by the structural income in 2007 (2010).

Third, following Naschold (2012) and Mukasa (2015) I implemented again the two-step procedure, but this time the first step was estimated by means of a fix effects model per household using the four years I have data for in the FLSS (2007, 2008, 2009 and 2010). This implies that now the coefficients of the explanatory variables are restricted to be the same for all the years. In the second step, again, a Kernel-weighted local polynomial smoothing regression of the structural income (i.e predicted income) in 2010 (2013) explained by the structural income in 2007 (2010) was estimated.

Fourth, using the method proposed by Zhou and Turvey (2015), I implemented the same two steps procedure, but in this case acknowledging that the current income in the first step may be simultaneously determined with the years of education of the household (represented by the years of education of the head of the household, average education of the workers or total education of the workers). I used as instruments the education level reached by the head of the household of the household where the head of the household grew up (presumably the education of the father or mother of the head of the household), the working status of the head of the household of the household where the head of the household grew up (presumably the working status of the father or mother of the head of the household in the past) and the self-reported poverty status of the household where the head of the household grew up.

Finally, I run a semi-parametric regression in the fashion of Kwak and Smith (2013), Gomez and Lopez (2013), You (2014) and Mukasa (2015), using as dependent variable the adult equivalent income as proportion of the poverty line in 2010 (2013), including in the parametric part of the regression the same controls than in the OLS model, and in the non-parametric part the adult equivalent income as a proportion of the poverty line in 2007 (2010) for capturing non-linearities on the dynamics of the income. In particular, I used Robinson's semiparametric regression estimator.

After estimating this set of models as the baseline, I jump into the correction of the aggregation problem of households in urban areas. Two problems should be addressed here. First, the aggregation of the exogenous variables. Second, the possible income underreport of the self-employed households. For solving the first problem, multiple papers have shown that theoretically different aggregations are possible. Yang and Yuying (2002) use the total education and total work experience of the members of the household, while Laszlo (2005) uses average education of members of the household 15+ years old. Besides these examples, most of the literature in this field uses the education and work experience of the head of the household when estimating the total income of the household. Here, I compare the results of the poverty trap estimations using these three measures of the education and work experience of the household when possible. For solving the second problem, the estimations of the poverty trap were repeated, but this time correcting for the potential underreport of income of the self-employed households.

For this purpose, I follow the Pissarides and Weber (1989) methodology. As explained before, trying to find to what extent self-employed households underreport their income possess a series of empirical problems, since the Engel curve is a stable relationship between the long-term income and the expenditure. The problem then becomes how to estimate the expenditure as a function of the long run income and the self-employed status of the head of the household. Most of the literature, including Hurst, Li and Pugsley (2014), Kim, Gibson and Chung (2017), Engstrom and Holmlund (2017), Johansson (2005) and Pissarides and Weber (1989) use an IV approach for estimating the effect of the long run income in expenditure. Those papers estimate an IV model of the current expenditure as left hand side variable using as right hand side variable the current income, but instrumenting such income with education. Conceptually this means that, first, education is highly correlated with the short-term income, and that the prediction of a regression between the current income and education can be interpreted as the long-term income. Second, education only affects expenditure through long run income but not through any other control included in the regression.

Hurst, Li and Pugsley (2015) and Engstrom and Hagen (2017) use as proxy for long run income a three years average of the current income and then run the model by OLS while Kim, Gibson and Chung (2015) uses a between effects model which itself averages the right hand and the left hand side variables in the model.

Finally, Kukk and Staehr (2014) note that the way in which a self-employed household is identified is critical in the results. Therefore, they consider two different definitions; the self-reported employment status of the head of the household and a definition based on the share of total household income coming from business-related income. The latter definition assumes a given threshold and ascribes a household as self-employed if the share of business related income in total income exceeds this threshold. I also investigate the importance of the choice of this threshold value.

Once the underreport rate for the income of the self-employed households is estimated, the household income is corrected by that rate and the poverty trap model is estimated again to see if such correction has any effect on the poverty trap result.

6. Results

6.1 Fedesarrollo Longitudinal Social Survey: 2007-2010

Figure 6 shows the result of a kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2010, on the same variable for the year 2007. The top left graph shows the result of the model estimated using an Epanechnikov kernel while the top right graph shows the results of the model using a Gaussian kernel. Both graphs on the top, use a rule of thumb to select the bandwidth and a second-degree polynomial approximation in the smoothing process. The bottom graphs use data driven, optimal bandwidth selection methods, a linear local approximation, an Epanechnikov kernel, and restricted the sample to households with an adult equivalent per capita income divided by the poverty line less than 15. In terms of the shape of the estimation all results are very similar: the shape of the estimated function is nonlinear and oscillating around the 45-degree line showing on some parts of the domain the S-shaped hypothesized by Carter and Barrett (2006). Every point where the estimated function crosses the 45-degree line can be considered a dynamic equilibrium of the adult equivalent per capita income per household. Therefore, after identifying them it should be identified whether such equilibrium is stable or unstable and whether it is statistically significant.

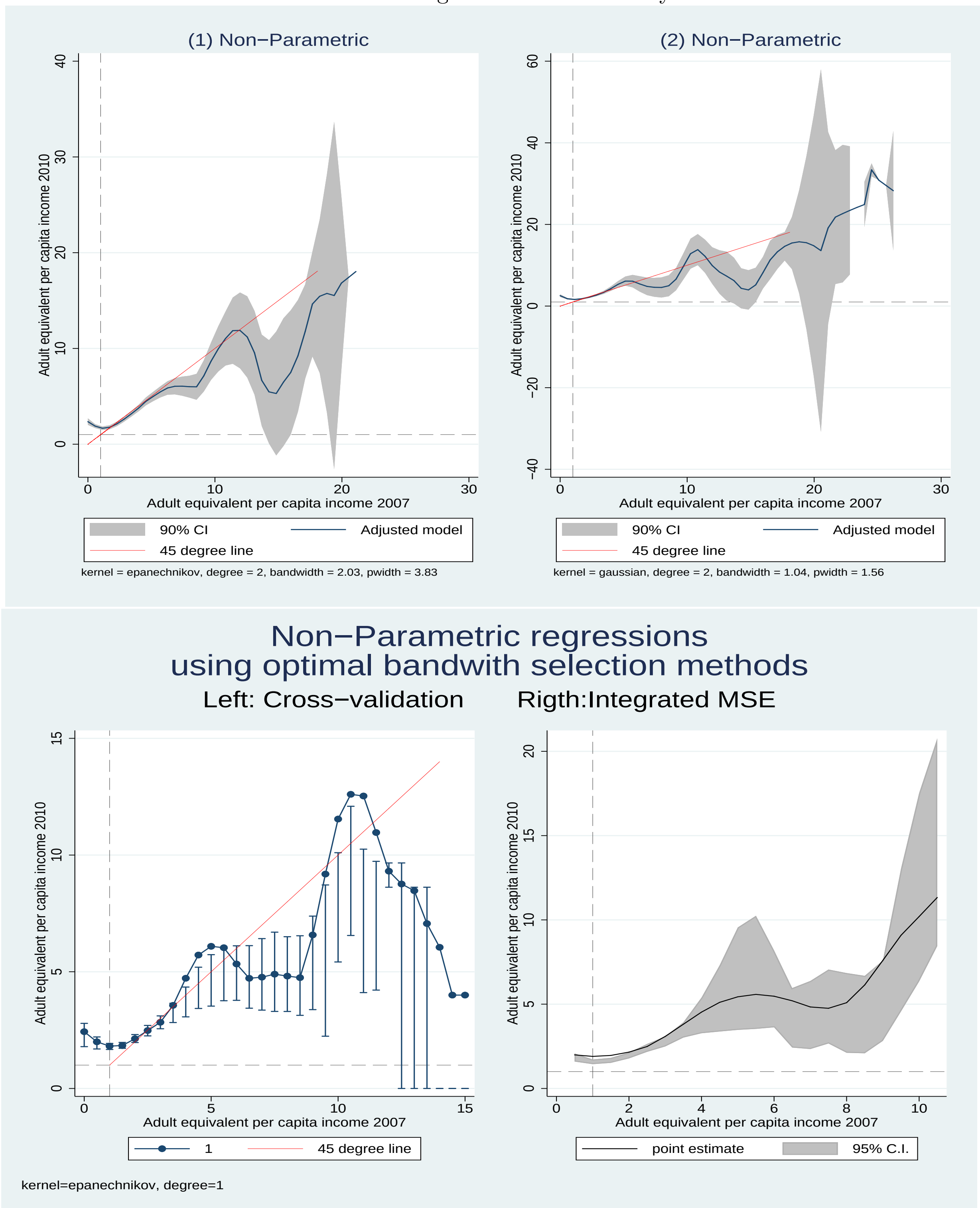
One equilibrium can be defined as statistically significant if the punctual estimation crosses the 45-degree line, while also the interval of the estimated function crosses the 45-degree line completely, before and after the crossing point of the punctual estimation and in different directions. In other words, for instance, an equilibrium would be significant if the interval containing it, changes from being completely above the 45-degree line to be completely below the 45-degree line.

Following such criteria, the top graph on the left shows only one stable equilibrium just above one (one being the dotted line) as the estimated interval never crosses completely above the 45-degree line. The figure on the top right shows instead, between 3 and 4 significant equilibria in an S-shaped function more in line with what was proposed by Carter and Barrett (2006). In both cases, the lower equilibrium crosses above 1 implying that despite the non-linearity of the household income dynamics, no evidence of a poverty trap is found. The bottom graphs test if said non-linearities are driven by outlayers in the top portion of the distribution, the choosing of bandwidths or the polynomial approximation. The result shows that despite the existence of less significant equilibria, the fitted models still show important non-linearities.

Figure 6

Non parametric regression of the Household per Capita Income Divided by the Poverty Line 2007 (Horizontal axis) vs. 2010 (Vertical axis)

FLSS: Fedesarrollo Longitudinal Social Survey. 2007-2010



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS). Note: 1/ Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2010, on the same variable for the year 2007. 2/ Top graphs: A rule of thumb were used to select the bandwidth and a second-degree polynomial was used in the smoothing process. The left graph was estimated using an Epanechnikov kernel while the graph on the right uses a Gaussian kernel. 3/ Bottom graphs: Optimal bandwidth selection methods, Epanechnikov kernel and a linear-degree polynomial were used in the smoothing process.The range of the variables was truncated at 15. The left graph uses cross validation bandwidth selector, while right graph uses integrated MSE. 4/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Table 2 implements the two-step procedure, in fashion of Adato, Carter and May (2006), Naschold (2012) and Zhou and Turvey (2015). Here, the table shows the results if the income generating function of the household is approximated using head of the household variables as proxy for inputs of the household. In particular, the model controls for age and age square as proxy for work experience, years of education as human capital input, the employment rate among the members of the household, calculated with the same formula used by the Colombian statistical office[2], the proportion of members of the household between 0 and 12 years old, the proportion of members of the household 62+ years old, the number of self-employed members of the household, using the self-reported status, and two city dummies for the cities of Bogotá and Cali.

Columns 1 and 2 run the model for the years 2010 and 2007 using OLS allowing for the coefficients in the income generating function to change between the two years. Column 3 estimates de model for the four years I have data (2007, 2008, 2009 and 2010) using fixed effects which in turn implies that the effect of the inputs in the income do not change on different years. Columns 4 and 5 repeat the OLS exercise for both years, but acknowledging the possible endogeneity of the variable education in respect to the income. Therefore, an instrument that is correlated only with education and correlated with the household income only through the education should be used. In this case, the survey asks for the education of the head of the household where the head of the household grew up, the working status of the head of the household where the head of the household grew up and the self-reported economic status of the household where the head of the household grew up.

[2]For being more precise with the working status variables, various definitions should be considered. First of all, the total population can be divided in working age population (people between 12 and 65 years old) and those who don't belong to the working age population (people below 12 years and older than 65 years). The working age population is divided among the occupied individuals (people who declare that they spent most part of the last week working, or people who declare that they spent most of the time during the last week working in his/her own business), the unemployed (people who claim that was looking for a job during the last four weeks and who was able to accept one in case of being offered some but didn't find one) and the inactive (people who declare that is permanent unable to work due to physical condition, people who declare that do not want to find a job or start a business, people who spend most of their time studying, people who haven't tried to find a job in the last four weeks, and those who are not available to take a job in case of being offered). The occupied fraction of the population can be divided in employee, sub-employed, independent worker (including employer) and housekeeping activities. The employment rate is defined as the quotient between the occupied population and the working age population.

In general, the results show that the variables have the expected sign, or in case of having the contrary sign, they are not significant. Age of the head of household have a positive effect on the adult equivalent per capita household income while its square is negative. In both cases, they are not significant. The years of education and employment rate of the members of the household have a positive and significant effect on the income. In the FE model education has a negative sign but it is not significant. The proportion of members between 0 and 12 years old have a negative and significant sign as they do not work and therefore, do not bring income home. The proportion of members 62+ years old is not significant as some of them may not produce any income at all, but some others are pensioners. Finally, the number of self-employed members are significant only in the FE model.

This set of estimations were repeated using as inputs the average age of the working members of the household and the average education of the working members of the household. Appendix 1 shows the result of such model. In general, the results are the same as for the model using the head of household variables. All the variables have the expected sign, but only average years of education, employment rate among household members and proportion of household members between 0 and 12 years old are consistently significant among estimation methods. Different from the models including head of the household variables, in this case in a couple of models the proportion of members 62+ years old have a significant negative effect on the adult equivalent per capita income. Appendix 2 repeats the same estimations but now using total age of the workers among household members and total education of the workers among household members. The results are basically the same, but now the total age of the workers seem to have a negative effect on the adult equivalent per capita income.

Table 2, Appendix 1 and Appendix 2 are different ways to estimate the first step in the two step approach used by Adato, Carter and May (2006), Naschold (2012) and Zhou and Turvey (2015). The second step is running a non-parametric regression between the predicted adult equivalent per capita income divided by the poverty line in 2010 and the prediction for the same variable in 2007. Figure 7 shows such regressions using a kernel-weighted local polynomial smoothing regression with a rectangular kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb. Row 1 shows the results using averages for the exogenous variables.

In this case, the results are somehow mixed and show the presence of zero or one equilibria as the long run income in 2010 tend to be always above the 45-degree line. In the cases where there is one equilibrium, it is above one, ruling out the presence of a poverty trap. Row 2 shows results using totals for the exogenous variables. The FE model shows two equilibria in an S-shaped function while the IV model shows one unstable equilibrium below 1. When using head of the household exogenous variables the FE model shows again an S-shaped estimated function with three equilibria two stable and one unstable. In any case, all of them are well above 1.

The practical use of the predicted income as a proxy for the long run income deserves further comment as it has many conceptual implications. First, I am using the covariates with their changes trough time. This, jointly with the fact that the OLS and IV models are run separately for each year, implies that the structural income may change due to changes in the covariates or changes in the estimated coefficients. Second, the FE model use all the 4 years of data for estimating the structural income. In this case, the estimated coefficient is the same for the four years, and as a consequence, all the variation in the structural income comes only from the changes in the covariates. Third, two of the most important variables in the models are either fixed or have a deterministic trend (age and education) when controlling for head of the household variables. Fourth, the education and age variables of some individuals (around 25% of the sample) present some inconsistencies on the reported data (decreasing age or education in some periods). Hence, a procedure for correcting this inconsistencies was applied, in an attempt to eliminate spurious variation in the exogenous variables that would translate into the structural income.

This characteristics are relevant, as most of the models based on Carter and Barrett (2006) allow covariates and coefficients to change over time, making the changes in structural income a reflection of the changes in covariates and coefficients. Nevertheless, some other questions may follow. For instance, what would be the effect on income for the households with fixed covariates through time?; or what would be the effect on total income for the population with fixed values of the covariates and coefficients through time?. In the last case, the structural income of the household would be the same across periods, and all the fluctuations in the observed income would be captured by the model error (or income shocks). Indeed, as argued by Sandoval (2019), to understand the full picture of the evolution of poverty is not enough to analyse only the evolution of the structural income, but said analysis should be complemented with an analysis of the income shocks (or error in the structural income model). The type of analysis assuming fixed covariates and interactions between structural income and income shocks are beyond the scope of this study but is of interest for future research.

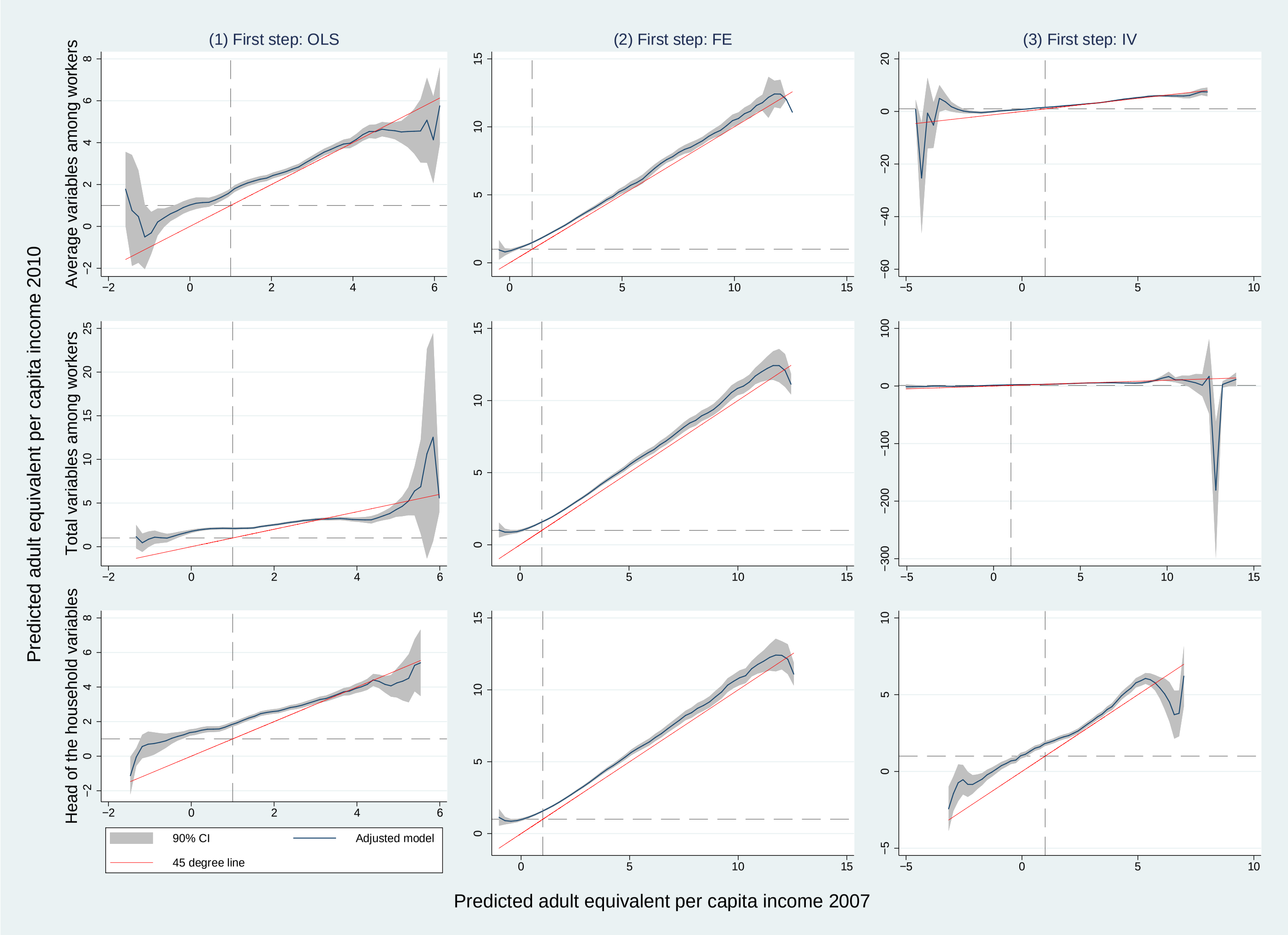
Table 2: Fedesarrollo Longitudinal Social Survey Two step results. 2007-2010. Dependent variable: Adult equivalent per capita income divided by the poverty line. Controls: Head of the Household Variables

	(1)	(2)	(3)	(4)	(5)
VARIABLES	OLS (2007)	OLS (2010)	FE	IV (2007)	IV (2010)
Age of head of the household	0.0153	0.0453	0.0061	0.0379	0.0759
	(0.036)	(0.056)	(0.046)	(0.039)	(0.058)
Age of head of the household 2	-0.0000	-0.0001	-0.0000	-0.0000	-0.0001
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Education of head of the household	0.1611***	0.2329***	-0.0239	0.3344***	0.4006***
	(0.017)	(0.025)	(0.041)	(0.034)	(0.049)
Employment rate among household members	3.2935***	2.1784***	1.6843***	3.3030***	2.0646***
	(0.292)	(0.431)	(0.209)	(0.313)	(0.446)
Proportion of members between 0 and 12 years old	-2.3532***	-2.2357***	-0.2528	-1.7377***	-1.5229*
	(0.472)	(0.782)	(0.551)	(0.520)	(0.835)
Proportion of members 62+ years old	-0.3699	0.1582	-0.0942	-0.2610	0.1865
	(0.368)	(0.513)	(0.424)	(0.397)	(0.530)
Number of self-employed household members	0.0700	0.0440	0.2982***	0.1305	0.0626
	(0.078)	(0.132)	(0.065)	(0.085)	(0.136)
Bogota	-0.1468	0.4076		-0.3915	0.1894
	(0.376)	(0.567)		(0.404)	(0.586)
Cali	-0.0800	0.3952		-0.3879	0.0952
	(0.377)	(0.568)		(0.407)	(0.589)
Constant	-1.7273	-3.6253**	0.8804	-4.6910***	-6.7433***
	(1.068)	(1.717)	(1.471)	(1.245)	(1.925)
Observations	684	684	2,739	677	677
R-squared	0.30	0.17	0.09	0.19	0.12
F	32.11	15.79	19.56	.	.
Number of id_hogar1			685		
Year FE			YES		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ The table presents the first step in the multi-step approach proposed by Adato, Carter and May (2006). Columns 1 and 2 present the results using OLS for estimating the first step for 2007 and 2010 separately. Column 3 show the results, estimating the first step model using all the years available (2007, 2008, 2009, 2010) through an FE data panel model. Columns 4 and 5 estimates the first step model using IV for the years 2007 and 2010 assuming that the education of the head of the household shows reverse causality with respect to the adult equivalent income divided by the poverty line. Education is instrumented with variables of the parents of the head of the household such as education of the head of the household where the head of the household grew up, the working status of the head of the household where the head of the household grew up and the self-reported economic status of the household where the head of the household grew up.

Figure 7: Fedesarrollo Longitudinal Social Survey. Second step of the two step method. 2007-2010



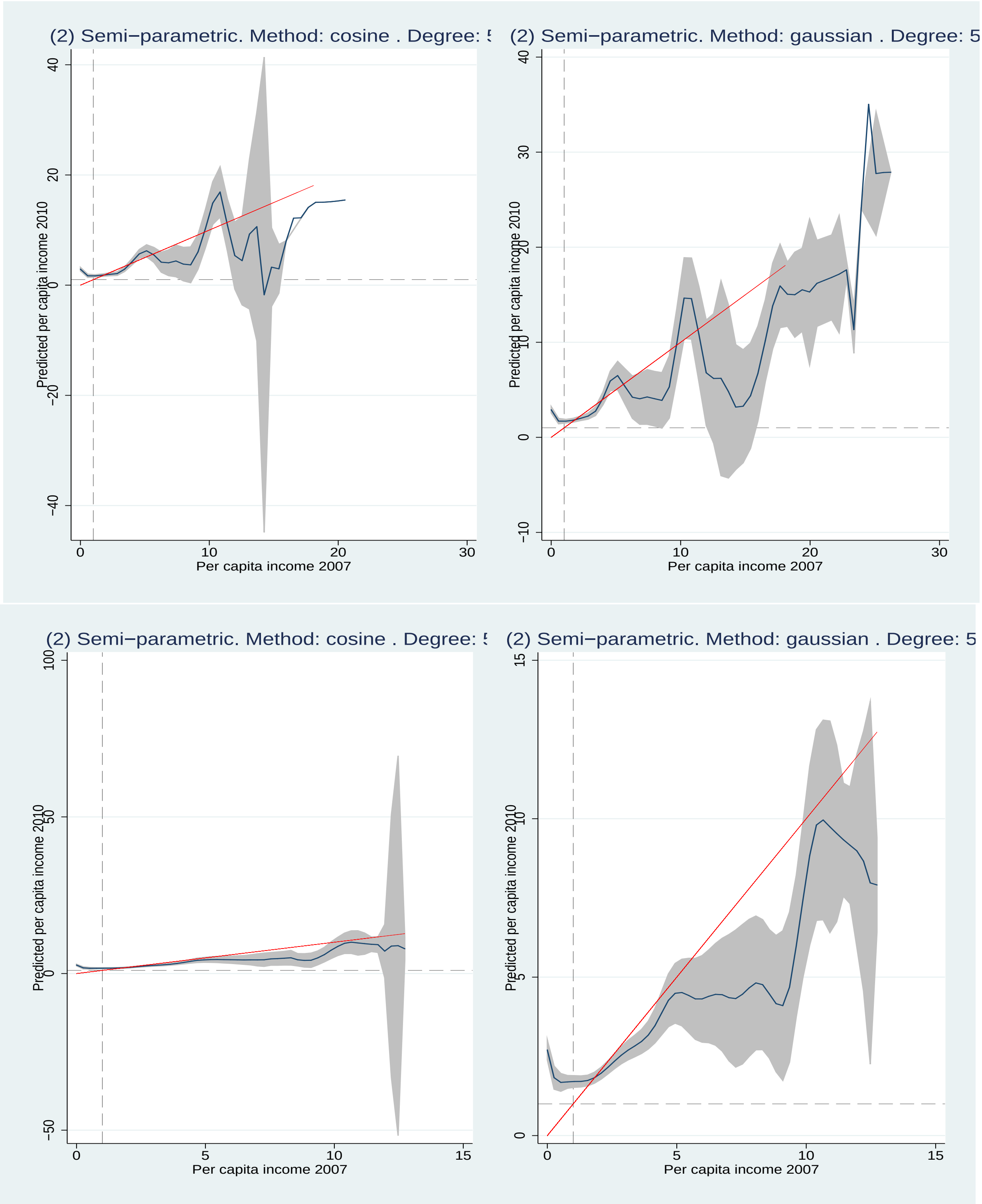
Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ The graph presents the second step in the multi-step approach proposed by Adato, Carter and May (2006). The vertical axis variable is the predicted adult equivalent per capita income divided by the poverty line in 2010 and the horizontal axis variable is the prediction for the same variable in 2007. 2/ The non parametric model was estimated using a kernel-weighted local polynomial smoothing regression with a rectangle kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb. 3/ Row 1 shows the results using averages among workers in the household for the exogenous variables. Row 2 shows results using totals among workers in the household for the exogenous variables. Row 3 shows the results when using head of the household variables as exogenous. 4/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line. 5/ Limiting the range of the variables at 15.

Appendix 3 shows the results of different specifications for estimating the poverty trap model using the Robinson's semiparametric estimator. Columns 1, 4 and 7 uses averages for exogenous variables, columns 2, 5 and 8 uses totals for exogenous variables. Columns 3, 6 and 9 uses head of the household controls. Different kernel methods were also used. Columns 1-3 uses Epanechnikov kernel while columns 4-6 uses Cosine kernel and columns 7-9 uses a Gaussian kernel. The usual variables are significant with the right sign: education, employment rate among working members and proportion of members between 0 and 12 years old. When using averages for the exogenous variables age and age square are also significant with a positive and negative effect respectively on the adult equivalent per capita income per household divided by the poverty line in 2010.

Figure 8 shows the non-parametric part of the model for totals on the exogenous variables and cosine and Gaussian kernels. The graphs at the top correspond to the columns 5 and 7 on appendix 3, while the graphs at the bottom restrict the sample to households with income below 15. The figures show the estimated adult equivalent per capita income per household divided by the poverty line in 2010 as a function of the observed adult equivalent per capita income per household divided by the poverty line in 2007 relating them through a non-parametric function. In the graphs at the top, there exist between 3 and 5 dynamic equilibria, as predicted by Carter and Barrett (2006). In the graphs at the bottom there is at most one significant equilibrium. As with the non-parametric specifications and two step specifications, in all cases the lowest equilibrium is above one, ruling out the existence of a poverty trap.

Figure 8

Predicted adult equivalent per capita income divided by the poverty line in 2010
Fedesarrollo Longitudinal Social Survey: Semi-parametric results. 2007-2010.



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ Variable in the vertical axis is the predicted adult equivalent per capita income in 2010, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2007. 3/ Top graphs uses all the range in all variables, bottom graphs truncated the range at 15. 4/ All graphs use a rule of thumb for choosing the bandwidth and a polynomial approximation of the 5th degree. The graphs on the left use a cosine kernel, while the graphs on the right use a gaussian kernel. 5/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

6.1.1 Income under-report for self-employed households

The purpose of this section is to determine whether household income reported in the Fedesarrollo Longitudinal Social Survey suffer from the income underreport of the self-employed households described by Pissarides and Weber (1989). Appendix 4 shows the results of the total income underreport estimation using different estimation methods (OLS, BE and IV) and different self-employed household definitions as suggested by Kukk and Staehr (2014).

The different self-employment definitions include many variations. Definition 1 implies that a household is self-employed if at least one of the adult members of the household report being self-employed. In this case a person is considered to be self-employed if he works as a pawn, he works on his own business, or he is an employer. In definition 2, I use the proportion of the income that is derived from business and the fees received from independent work as a measure of the degree on how self-employed the household is. Right after such a measure is created I define a threshold above which a given household is considered self-employed.

Kukk and Staehr (2014) claim that the results may vary importantly depending on whether the household can be permanently classified as self-employed. In other words, when using longitudinal data, two distinctions should be made: first a household may be considered self-employed only when it is self-employed in every single period. In the same way, a household is not self-employed when it is not self-employed in any of the observed periods. This definition implies that the households changing self-employed status through time should be ignored, restricting the sample size. Second, in contraposition to this, if the household is considered self-employed period by period individually, the sample is unrestricted.

All the possible combinations for such models were estimated. Appendix 4 shows some of this models, while the others are available on request. In all cases the right-hand side variable is the household food expenditure. The four years in the panel were used, leading to a potential of 2740 observations (685 households followed in a period of 4 years).

Columns 1 and 2 show the results for the OLS models. In this case a three-year average income is used as a proxy for long run income. Column 1 uses the unrestricted self-employed definition 2, with a threshold of 25%, i.e a household is considered to be self-employed if more than 25% of its income is derived from business and fees paid for independent work. Column 2 uses the same specification, but now following the restricted self-employed definition. In both cases being a married couple, the size of the household, the years of education of the head of the household and the long run income have a positive effect on the household food expenditure. In the first case, the estimated underreport rate is 38.4%. In the second case, it is 70%, but with a much smaller sample.

Column 3 repeats the exercise but this time using the between effects model. This type of model uses averages of the variables along the time dimension, mitigating a little the endogeneity between household income and food expenditure. The significance and direction of the effects are similar to those on the OLS models. In this case the estimated underreport rate is 68.1%, using the self-employed definition 1 for the unrestricted sample.

Columns 4-6 repeat the exercise but now using the IV estimation method. In this case, the household income is instrumented with variables that are supposed to be correlated directly with household current income, and with food expenditure only through long run income. The literature typically uses education of the head of the household as a determinant of long run income. This survey in particular has other information like the education of the head of the household where the head of the household was raised, the working status of the head of the household where the head of the household was raised and the self-reported economic status of the household where the head of the household was raised. Using such variables as instruments, I found that only household income and self-employment status are significant as explanatory variables of the food expenditure. In those models the underreport rate varies from 17.3% to 31.6%.

In all models the self-employment status is significant.

After finding the correction factor on the income, the next step is to proceed to correct the income of the self-employed households inflating their income by the corresponding estimated rate and then re estimate the poverty trap models. Appendix 4a and 4b does exactly this, combining the different methods for estimating the underreport rate (OLS. IV, BE) with non-parametric and semiparametric estimation methods for the poverty traps. Those two graphs use the long (unrestricted) sample. In this case, the general result of the poverty trap models holds: they show a highly nonlinear behaviour of the income dynamics around the 45-degree line with multiple equilibria. Sometimes with 0, 2 or 3 significant equilibria.

Appendixes 4c and 4d repeat the exercise but using the short (restricted) sample. In this case the general shape of the function still holds, but the estimation loses power due to the drastic reduction on the sample size.

Finally, it is important to acknowledge that I am correcting the income variable, only with the punctual estimation of the income underreport rate for the self-employed households. As such rate is itself an estimation, its own standard error should also be estimated and in turn, when making the poverty trap estimation, the standard error of both estimations should be accumulated. In other words, appendixes 4a-4d shows only the smallest possible interval after correcting for income underreport of the self-employed, and even in such case, the intervals for the poverty trap estimations almost always contain the 45-degree line. If I were correcting the intervals for taking into account the variance of the prediction of the underreport rate, the amplitude of this interval would almost certainly always contain the 45-degree line, leading to the conclusion that there is not any statistically significant equilibrium.

6.2 Colombian Longitudinal Survey by Universidad los Andes: 2010-2013

This section deals with the estimation of poverty traps in urban areas in Colombia, but using a different survey with more coverage in terms of sample size and representativeness. Under this sample the estimations can be made for the 14 main Colombian cities and other urban areas.

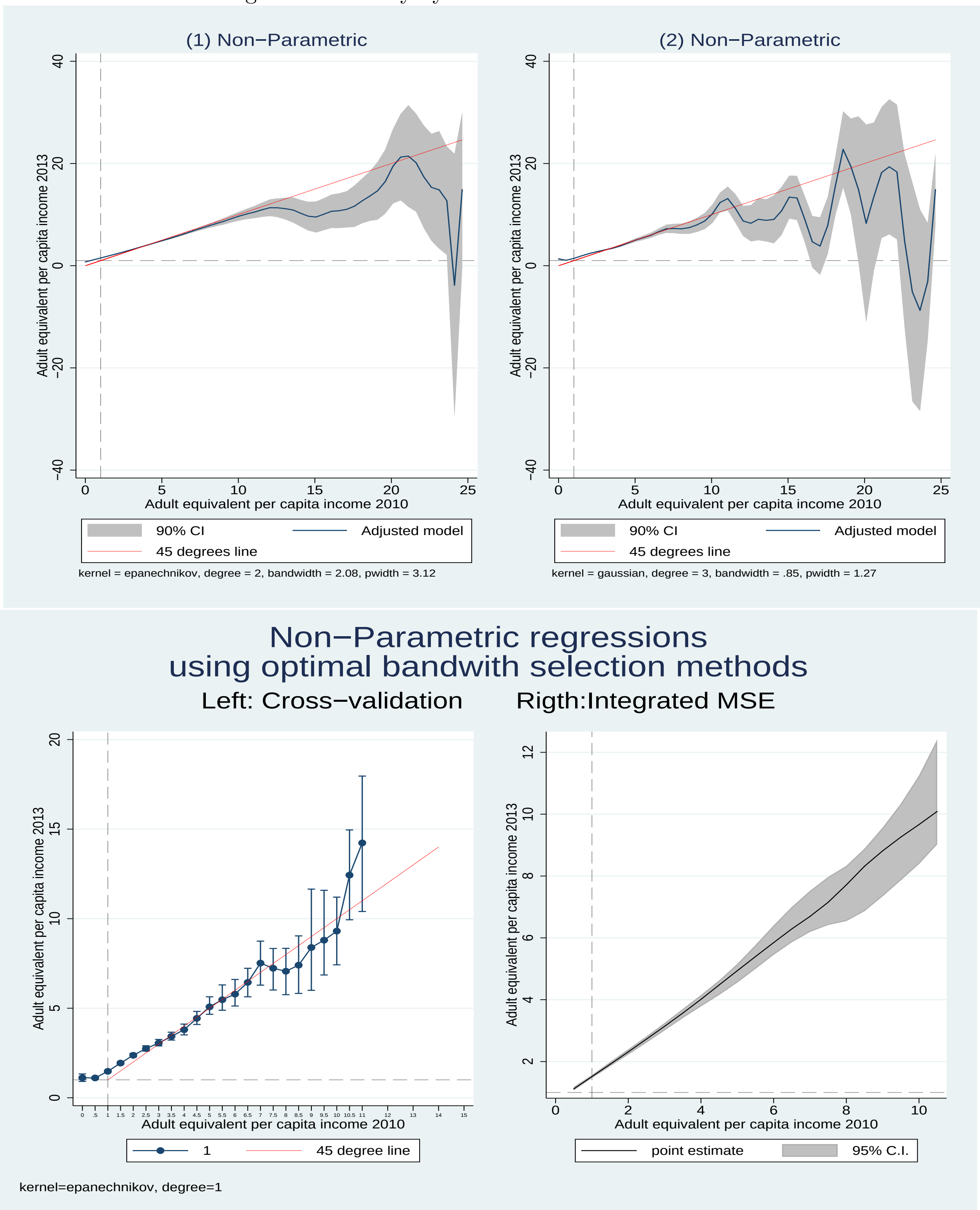
Figure 9 shows the results of repeating the non-parametric estimation conducted in figure 6, under the same specification, for all the graphs, but in this case using as dependent variable the adult equivalent per capita income per household divided by the poverty line in 2013 and as independent variable the adult equivalent per capita income per household divided by the poverty line in 2010 for all urban areas in Colombia, including 14 main cities and other urban areas.

At first sight, the results seem to be very similar of those of the Fedesarrollo Longitudinal Social Survey. The estimated dynamics of the adult equivalent per capita income per household fluctuate around the 45-degree line. When using the Epanechnikov kernel there seems to be only one equilibrium (top left), while when using the Gaussian kernel there seems to be two or three equilibria (top right). In any case, none of the equilibria are below one, implying that there is not a poverty trap despite existing a non-linear behaviour on the income dynamics and an S-shape in such dynamics as predicted by Carter and Barrett (2006). Nevertheless, when repeating the exercise restricting the sample to households with an adult equivalent per capita income divided by the poverty line below 15 and data driven bandwidth choosing methods, the result indicates that the estimated function is mostly linear. This is evidence in favour of the hypothesis that the results for the graphs at the top, are mainly driven by outliers or non-linearities in the highest percentiles of the distribution. Indeed, when using this database, the households with higher income in 2010 (households with adult equivalent per capita income in 2010 of 15 and above) have lost income in 2013 as most of the estimated function above 15 is below the 45-degree line.

Figure 9

Non parametric regression of the Household per Capita Income Divided by the Poverty Line 2010 (Horizontal axis) vs. 2013 (Vertical axis)

CLSA: Colombian Longitudinal Survey by Universidad los Andes. All urban areas. 2010-2013



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes (CLSA). Note: 1/ Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2013, on the same variable for the year 2010. 2/ Top graphs: A rule of thumb were used to select the bandwidth and a second-degree polynomial was used in the smoothing process. The left graph was estimated using an Epanechnikov kernel while the graph on the right uses a Gaussian kernel. 3/ Bottom graphs: Optimal bandwidth selection methods, Epanechnikov kernel and a linear-degree polynomial were used in the smoothing process. The range of the variables was truncated at 15. The left graph uses cross validation bandwidth selector, while right graph uses integrated MSE. 4/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Appendixes 6A and 7A repeat the non-parametric estimation procedure, but now separating the results between 14 main cities as in the Appendix 6A and other urban areas as in Appendix 7A. It is not strikingly surprising that the behaviour of the income dynamics among those two areas is very different. On the one hand, the 14 main cities estimation shows that there is a stable lower dynamic equilibrium just above 1, a second nonstable equilibrium around 12 and a third dynamic stable equilibrium around 14. On the other hand, the estimation for the other urban areas shows consistently 0 or at most 1 statistically significant dynamic equilibrium, implying that the social mobility on this other urban areas in Colombia is almost null. In other words, statistically the income of the households in those areas remain the same in both years, as the 45-degree line is contained by the estimated interval in almost all the range of the function. Therefore, the result in figure 9 is mainly driven by the income dynamics on the 14 main cities, implying that those two groups of cities should be analysed separately.

Table 3 repeats the set of estimations of table 2: the two-step approach using the 2010 and 2013 data for all urban areas. In this case the Colombian Longitudinal Survey by Universidad de los Andes do not ask for labour variables to all household members on both years, which in turn implies that different ways to aggregate the exogenous variables are not possible. Instead, the survey only asks for such variables for the head of the household and its partner making only possible to compare among the results when using only head of the household variables as controls, or head of the household and partner variables as controls.

Columns 1,2,5, 7 and 8 show the results when controlling only for head of the household variables, while the remainder columns show the results when controlling for head of the household and partner variables. As in table 2, IV models were estimated for acknowledging the endogeneity of education with respect to income. In this case education of the head of the household is instrumented with education of the father of the head of the household, education of the mother of the head of the household, working status of the father of the head of the household and working status of the mother of the head of the household. When controlling also for partner variables, the corresponding instruments were used.

The results are similar to those of table 2. Education of the head of the household and its partner is always significant and positive on the adult equivalent per capita income for all specifications except FE, where it is reasonable to assume that the education effect is being primarily captured by the household fixed effect. A consistently negative effect of the proportion of members between 0 and 12 years old on the adult equivalent per capita income was also found.

Figure 10 shows the results after applying the two step procedure, adjusting the kernel-weighted local polynomial smoothing regression with a rectangle kernel and a 4th degree smoothing polynomial on the predictions of the adult equivalent income in both years. For the models where the first step was estimated using OLS or IV the results show a non-stable equilibrium above one implying that either the income tends to zero or it just grow indefinitely. In the case of the FE models, one or three equilibria are found in an S-shaped way. In any case, none of the equilibria are below 1 ruling out the possibility of a poverty trap.

Appendixes 6B and 7B show the results of the estimation of those two step models for the 14 main cities and other urban areas respectively. The results remain unchanged: education of both partners is the most important variable when explaining the adult equivalent adjusted per capita income, followed by the proportion of household members between 0 and 12 years old.

Appendixes 6C and 7C show the graphical result of the second step in the two step procedure. In this case, the results are quite mixed for the 14 main cities as some models show one stable equilibrium, some others one unstable equilibrium and some others zero equilibrium. The results for the other urban areas are equally mixed: some of them have an S-shape with multiple equilibria while some others show one stable equilibria and in other cases the interval of the estimated function covers the 45-degree line over all the range.

Table 3: Colombian Longitudinal Survey by Universidad los Andes Two step results. 2010-2013. All urban areas.

Dependent variable: Adult equivalent per capita income divided by the poverty line

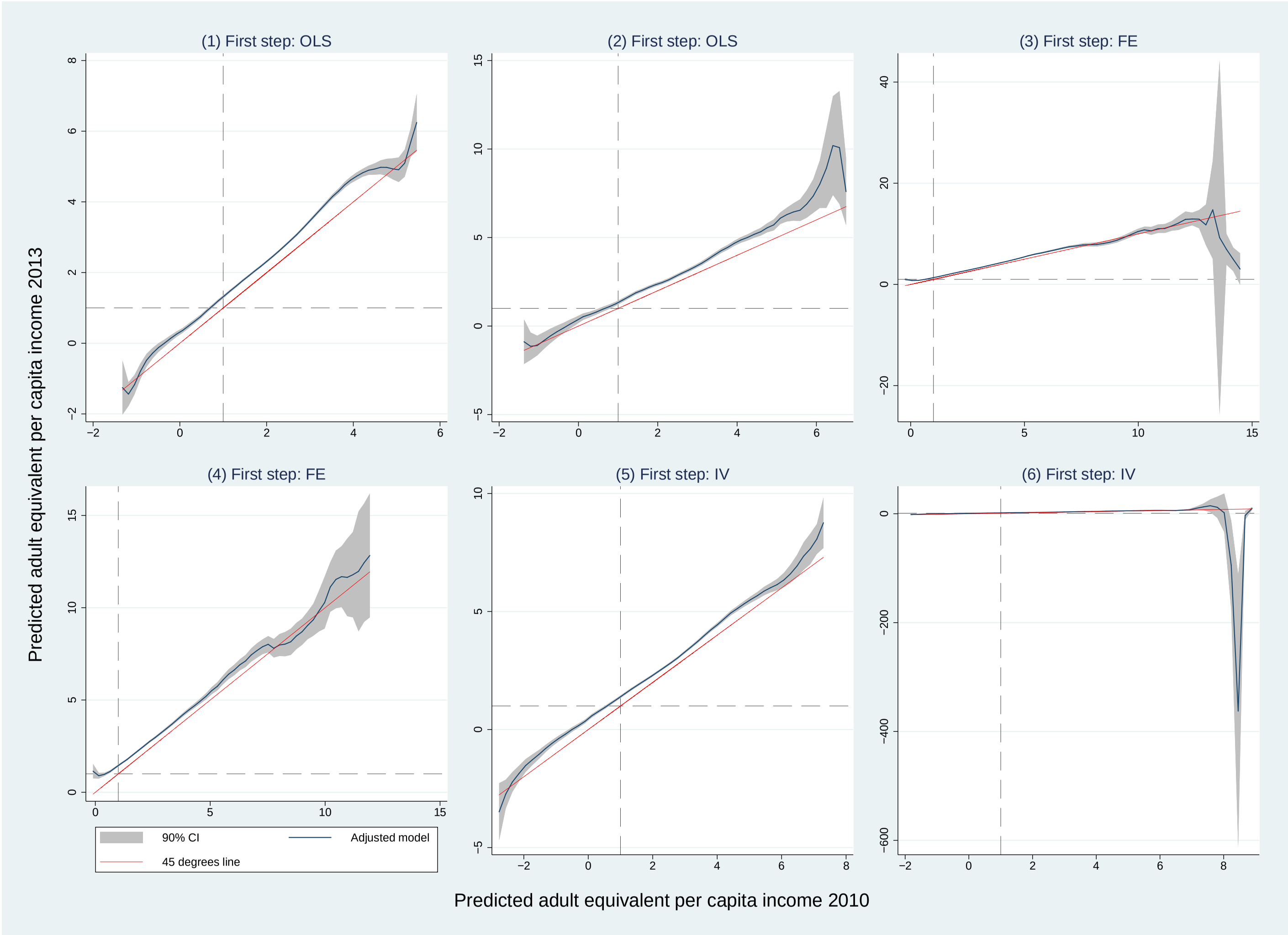
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	OLS (2010)	OLS (2013)	OLS (2010)	OLS (2013)	FE	FE	IV (2010)	IV (2013)	IV (2010)	IV (2013)
Age of head of the household	-0.0029	-0.0033	0.1010	0.1438	-0.1546	-0.3308	0.0089	-0.0178	0.1010	0.0708
	(0.021)	(0.033)	(0.062)	(0.108)	(0.131)	(0.332)	(0.029)	(0.045)	(0.083)	(0.141)
Age of head of the household 2	0.0003	0.0004	-0.0009	-0.0012	0.0026***	0.0024	0.0003	0.0007	-0.0008	-0.0004
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
Education of head of the household	0.1807***	0.2106***	0.1147***	0.1246***	0.0145	0.0470	0.3188***	0.3383***	0.2150***	0.1769**
	(0.008)	(0.010)	(0.021)	(0.031)	(0.028)	(0.065)	(0.023)	(0.031)	(0.056)	(0.086)
Self-employed head of the household	0.0011	0.0492	-0.0651	0.2458	-0.0832	-0.1718	0.1270	0.2454*	-0.2085	0.2773
	(0.069)	(0.095)	(0.153)	(0.232)	(0.073)	(0.161)	(0.092)	(0.136)	(0.188)	(0.290)
Age of head of household's partner			-0.1229**	-0.1445		-0.1492			-0.0409	-0.0291
			(0.061)	(0.103)		(0.215)			(0.079)	(0.132)
Age of head of household's partner 2			0.0016**	0.0017		0.0019			0.0006	0.0004
			(0.001)	(0.001)		(0.002)			(0.001)	(0.001)
Education of head of household's partner			0.1498***	0.1981***		-0.0337			0.1633***	0.1774*
			(0.021)	(0.034)		(0.044)			(0.051)	(0.091)
Self-employed head of household's partner			-0.0633	0.0040		-0.1202			-0.2096	-0.1397
			(0.152)	(0.233)		(0.154)			(0.199)	(0.318)
Proportion of members between 0 and 12 years old	-1.3599***	-1.7953***	-1.2172***	-1.8979***	-0.5609**	-0.7595	-1.4302***	-1.9232***	-1.8024***	-2.9357***
	(0.181)	(0.265)	(0.433)	(0.692)	(0.245)	(0.606)	(0.238)	(0.337)	(0.563)	(0.864)
Proportion of members 62+ years old	-0.3987	-0.0375	-1.7845*	-0.6136	0.1284	-0.6922	0.2495	0.6946	-1.0168	-0.8541
	(0.267)	(0.339)	(0.935)	(1.235)	(0.331)	(0.964)	(0.531)	(0.517)	(2.171)	(2.008)
Constant	-0.4123	-0.4072	-0.9220	-3.0026	3.3825	14.9982	-2.6161***	-2.3659**	-4.0178**	-3.8571
	(0.972)	(2.380)	(2.180)	(3.698)	(5.526)	(12.375)	(0.684)	(1.110)	(1.741)	(3.262)
Observations	2,718	2,718	666	666	5,436	1,332	1,860	1,860	404	404
R-squared	0.28	0.26	0.37	0.30	0.10	0.08	0.25	0.24	0.41	0.32
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
F	15.26	9.923	5.082	3.304	3.778	1.768	.	.	.	.
Number of llave_pers_const					2,887	713				
Year FE					YES	YES				

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ The table presents the first step in the multi-step approach proposed by Adato, Carter and May (2006). Columns 1 and 2 present the results using OLS for estimating the first step for 2010 and 2013 separately. Column 5 show the results, estimating the first step model using all the years available as a pool (2010 and 2013) through an FE data panel model. Columns 7 and 8 estimates the first step model using IV for the years 2010 and 2013 assuming that the education of the head of the household shows reversal causality in respect to the adult equivalent income divided by the poverty line. Education is instrumented with variables of the parents of the head of the household. The other columns includes as additional controls variables of the spouse of the head of the household.

Figure 10: Colombian Longitudinal Survey by Universidad los Andes. Second step of the two step method. 2010-2013. All urban areas

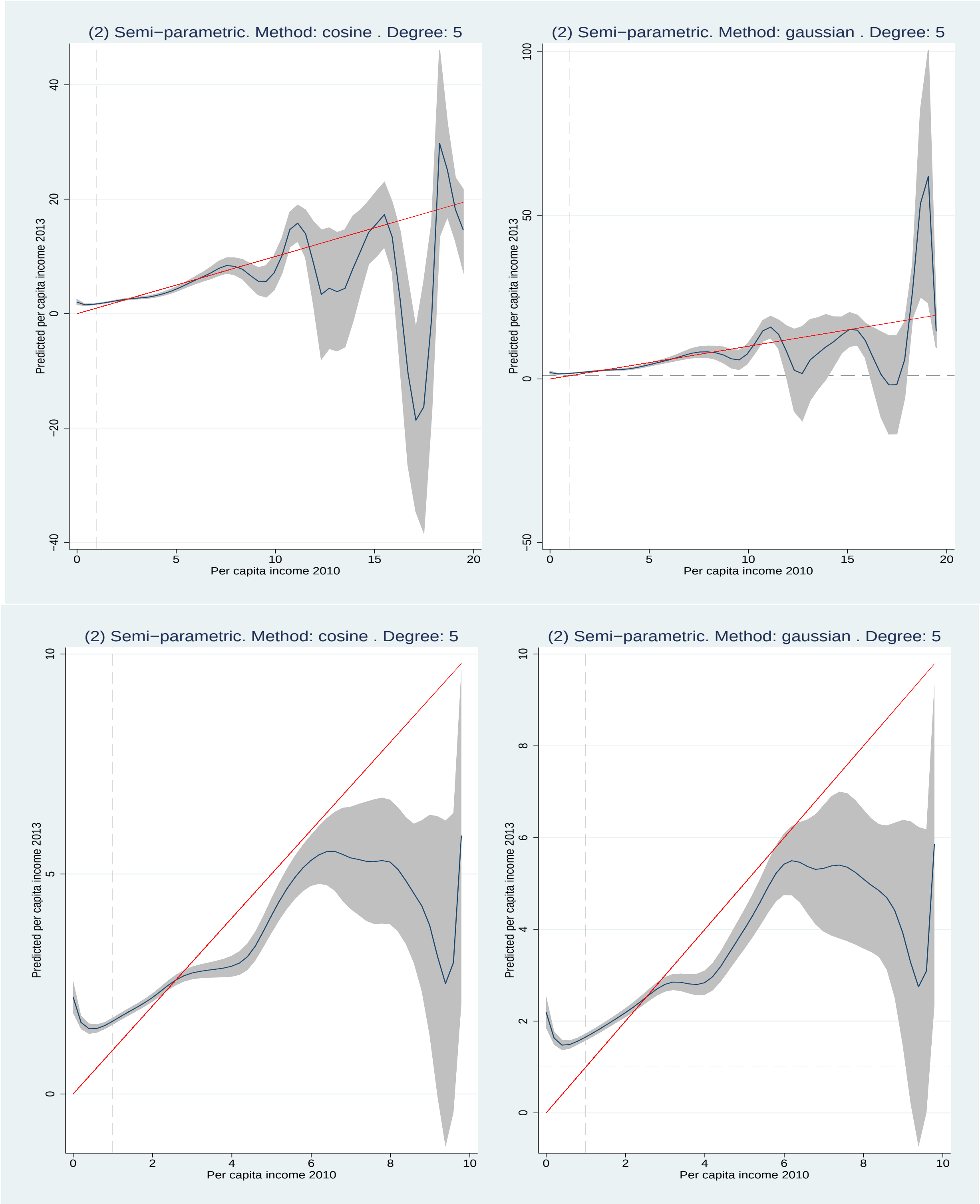


Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ The graph presents the second step in the multi-step approach proposed by Adato, Carter and May (2006). The vertical axis variable is the predicted adult equivalent per capita income divided by the poverty line in 2013 and the horizontal axis variable is the prediction for the same variable in 2010. 2/ The non parametric model was estimated using a kernel-weighted local polynomial smoothing regression with a rectangle kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb. 3/ Graphs 1 and 2 use the OLS first step results, while graphs 3 and 4 uses the FE first step results and graphs 5 and 6 uses the first step IV results. Odd graphs uses only head of the household controls, while even graphs uses head of the household and its spouse controls. 4/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line. 5/ The range of the variables was truncated at 15

Figure 11 shows the result of the semi-parametric model for all the urban areas. In this case the adult equivalent income presents multiple equilibria and an S-shape. As usual, none of the equilibria are below one ruling out the existence of a poverty trap. Appendix 6D and 7D repeat the estimations separately for 14 main cities and other urban areas. While the result seems to be consistent with those in figure 11 for the 14 main cities, there seems to be not enough statistical power to identify any effect in the case of the other urban areas.

Figure 11

Predicted adult equivalent per capita income divided by the poverty line in 2013
Colombian Longitudinal Survey by Universidad los Andes: Semi-parametric results. 2010-2013.



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ Variable in the vertical axis is the predicted adult equivalent per capita income in 2013, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2010. 3/ Top graphs uses all the range in all variables, bottom graphs limited the range at 15. 4/ All graphs use a rule of thumb for choosing the bandwidth and a polynomial approximation of the 5th degree. The graphs on the left use a cosine kernel, while the graphs on the right use a gaussian kernel. 5/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

6.3 Robustness checks

Various robustness checks were conducted on the Fedesarrollo Longitudinal Social Survey results. First, most of the estimations were redone using adult equivalent per capita expenditure, instead of income. In all of those cases the results tend to find only one stable equilibrium above 1 or non-equilibria at all, showing always an increase in the expenditure in 2010 in respect to 2007. As in previous literature, an Arellano-Bond and Blundell-Bover panel data model was estimated. In these cases, the results show the existence of a single stable equilibrium above 1.

For being sure that in the case of finding multiple equilibria the result is not being driven by outliers, the estimations were repeated dropping out 5% of the outliers (2.5% on every side of the distribution). In this case the shape of the estimated income dynamics do not change but its significance does. For the non-parametric and semiparametric estimates, multiple equilibria are found, but now being not significant. The parametric results remain very similar after correcting for outliers.

The models were also re estimated using different kernel functions, different degrees of the smoothing function and different bandwidths in the non-parametric and semiparametric steps. The general shape of the estimated function seems to be robust to the kernel function (epanechnikov, biweight, cosine, gaussian, parzen, rectangle and triangle) and the degree of the smoothing function. Nevertheless, the number of significant equilibria are not robust at the degree of the smoothing functions when they are greater than 2, neither to the kernel function. A parameter that completely changes the shape and significance of the estimated equilibria is the bandwidth selected for the estimation of the non-parametric function. I allowed this to be selected by a rule of thumb in the models shown here, but if a different bandwidth is selected probably the shape and significance of the estimated function will change completely.

In respect to the income underreport for the self-employed, all the specifications used in the literature were estimated. In around 40% of the models, it was found some evidence of self-employed household underreport on the income, while in the other models the self-employed status was not significant and sometimes had the opposite sign.

Originally, the Fedesarrollo Longitudinal Social Survey follows a panel of 685 households through a period of 4 years (2007 to 2010) across 3 cities. In 2008 the sample was increased to cover 14 cities, so when using such sample, I had data for 2.066 households across 3 years. The reason for not using such sample in the main estimations is that for being able to capture long run dynamics of income is much preferred to have observations as far as possible on time. Despite this, some authors have estimated such poverty trap models over shorter spans of time, ignoring the fact that it should be procured to capture changes in the structural income on the long run. Nevertheless, I repeat the estimation of the models for this three-year sample. In this case the results are better than with the 4-year sample. The non-parametric, the parametric and the semi-parametric models show an S-shaped dynamic of the income, in most of the cases with three significant equilibria.

7. Conclusions

This paper has tested the existence of poverty traps in Colombian urban areas using the only two household data panels existing for the country, covering the years 2007-2010 and 2010-2013. The main result is that for the 14 main cities no poverty trap of the type hypothesized by Carter and Barrett (2006) exists. For the other urban areas in Colombia the evidence is mixed.

This is the first paper, to the best of my knowledge, to test the existence of poverty traps in urban areas using a reliable measure of the total income of the household. This implied that several practical issues should be addressed. First, the poverty asset based approach works well in rural contexts where most of the income is derived from rural assets and there exist detailed information on such assets. As a consequence the variation of the income explained by the productive assets tends to be high, typically between 20% and 55%. In this paper, I found that the part of income that can be explained by assets varies between 14% and 37% for the urban areas. Second, there is an aggregation problem. Different from rural areas, in urban areas, members of the household typically work on different sectors of the economy, therefore, it is not reliable to assume that all the members of the household have the same income generating function as farmer workers. In this case, different aggregations for the exogenous variables were controlled for: head of the household variables, average variables for the workers and total variables for the workers. Those different aggregations do not change the substance of the results.

Another problem related with the origin of the household income was first pointed out by Pissarides and Weber (1998). They noted that self-employed households tend to underreport their income, which is something that may affect the poverty trap results. I found that even if I correct for such self-employed household income underreport, the main results do not change. There is no evidence of a poverty trap.

Something that affects the shape of the estimated income dynamics is the econometric method. The main hypothesis of Carter and Barrett (2006) is that if there exist a poverty trap, the long run income dynamics should behave in an S-shaped way, and have multiple equilibria: two of them stable and one non-stable. In the poverty trap case the lowest stable equilibrium is supposed to be below the monetary poverty line and the highest stable equilibrium above the monetary poverty line. When using more flexible estimation methods, like non-parametric or semi-parametric it is typical to find two or more significant equilibria whit an S-shape like the one hypothesised by Carter and Barrett (2006). When using two step methods, which combines in a first step a parametric method and in the second step a non-parametric method, the results consistently show one or at most two-significant equilibria. Under any estimation method used and independently of the number of equilibria found, always the lowest stable equilibrium is above 1, ruling out the existence of a poverty trap.

Other robust checks were made. Mainly to different specifications of the non-parametric and semiparametric parts (kernel function, bandwidth and degree of the smoothing function). Those changes affect the significance of the equilibria found, but not the general shape of the estimated function of the income dynamics.

Despite not having found a poverty trap, this paper opens the question under which conditions the household per capita income would converge to a low or a high equilibria. In other words, this paper diagnosed that only under certain flexible estimation methods there exist a multiple equilibrium on the income dynamics of the household. Nevertheless, the theoretical model resulting in such multiple equilibrium is yet to be found. A model like that would allow to answer to what extent exogenous income shocks have the power to determine to which equilibrium the income of a given household converges. The most interesting case would be if a poverty trap is found and an S-shaped income dynamics function with a multiple equilibrium is estimated. In such case a fourth-generation model of poverty measurement may be estimated, allowing to distinguish the structural foundations of poverty and giving a glance at the long-term persistence of structural poverty.

8.

References

[1] Adato, Michelle, Michael Carter and Julian May. Exploring poverty traps and social exclusion in South Africa using qualitative and quantitative data. *The Journal of Development Studies.* 42(2): 226-247. 2006

[2] Antman, Francisca and David McKenzie. Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. *The Journal of Development Studies.* 43(6): 1057-1083. 2007

[3] Arunachalama, Raj and Ajay Shenoyb. Poverty traps, convergence, and the dynamics of household income. *Journal of Development Economics.* Volume 126: 215-230. 2017

[4] Azarides, Costas and John Stachurski. Poverty traps. In Philippe Aghion and Steven Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1A. 2005

[5] Barrett, Christopher, Paswel Marenya, John Mcpeak , Bart Minten ,Festus Murithi , Willis Oluoch-Kosura, Frank Place , Jean Claude Randrianarisoa , Jhon Rasambainarivo and Justine Wangila. Welfare dynamics in rural Kenya and Madagascar. *The Journal of Development Studies.* 42(2): 248-277. 2006

[6] Barrett, Christopher and Michael Carter. The Economics of Poverty Traps and Persistent Poverty: Empirical and Policy Implications. *The Journal of Development Studies.* 49(7): 976-990. 2013

[7] Banerjee, Abhijit and Andrew Newman. Occupational Choice and the Process of Development. *Journal of Political Economy*, 101(2): 274–298. 1993

[8] Bourguignon, Francois and Satya R. Chakravarty. The Measurement of Multidimensional Poverty. *Journal of Economic Inequality.* 1(1): 25-49. 2003

[9] Bourguignon, Francois and Christian Morrison. Inequality among World Citizens: 1820-1992. *American Economic Review*. 92(4):727–744. 2002

[10] Calvo, Cesar and Stefan Dercon. Chronic Poverty and All That: The Measurement of Poverty over Time. CSAE Working Paper Series, Oxford University. 2007(4), 2007

[11] Carter, Michael and Christopher Barrett. The economics of poverty traps and persistent poverty: An asset-based approach. The *Journal of Development Studies*. 42(2): 178-199. 2006

[12] Carter, Michael and Travis Lybbert. Consumption versus asset smoothing: testing the implications of poverty trap theory in Burkina Faso. *Journal of Development Economics*. 99(2012): 255–264. 2012

[13] Dasgupta, Partha and Debraj Ray. Inequality as a Determinant of Malnutrition and Unemployment: Theory. *The Economic Journal*. 96(384): 1011–1034. 1986

[14] Dutta, Swati. Identifying Single or Multiple Poverty Trap: An Application to Indian Household Panel Data. *Social Indicators Research*. 120(1): 157–179. 2015

[15] Engstrom, Per and Johannes Hagen. Income underreporting among the self-employed: A permanent income approach. *European Economic Review*. 92(2017): 92-109. 2017

[16] FEDESARROLLO. Ficha técnica: Encuesta Social Longitudinal de Fedesarrollo 2010. Fundación para la Educación Superior y el Desarrollo, Bogotá, Colombia. 2010

[17] Filmer, Deon, and Lant H. Pritchett. Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. *Demography*. 38(1): 115 - 32. 2001

[18] Foster, James, Joel Greer, and Erik Thorbecke. A Class of Decomposable Poverty Measures. *Econometrica*. 52(3): 761-66. 1984

[19] Galor, Oded and Joseph Zeira. Income Distribution and Macroeconomics. *Review of Economic Studies*. 60(1): 35–52. 1993

[20] Giesbert, Lena and Kati Schindler. Assets, Shocks, and Poverty Traps in Rural Mozambique. *World Development*. 40(8): 1594-1609. 2012

[21] Gradin, Carlos, Coral. del Rio and Olga Canto. Measuring Poverty Accounting For Time. *Review of Income and Wealth*, 58(2): 330-354. 2012

[22] Gomez, Javier and Alejandro Lopez. Poverty dynamics in rural Mexico: What does the future hold?. *Ensayos Revista de Economía*. 32(2): 55-74. 2013

[23] Hien, Vu. Income Distribution and Poverty Traps in Vietnam. Mimeo. CIFREM, University of Trento. 2011

[24] Hurst, Erik, Geng Li, and Benjamin Pugsley. Are household surveys like tax forms?. Evidence from income underreporting of the self-employed. *The Review of Economics and Statistics*. 96(1): 19–33. 2014

[25] Jalan, Jyotsna and Martin Ravallion. Geographic poverty traps? A micro model of consumption growth in rural China. *Journal of Applied Econometrics*. 17(4): 329-346. 2002

[26] Johansson, Edvard. An Estimate of Self-Employment Income Underreporting in Finland. *Nordic Journal of Political Economy*. 31(2005): 99-110. 2005

[27] Kai-Yuen, Tsui. Multidimensional Poverty Indices. *Social Choice and Welfare*. 19(1): 69-93. 2002

[28] Kim, Bonggeun, John Gibson and Chul Chung. Using panel data to estimate income under-reporting by the self-employed. *The Manchester School*. 85(1): 41–64. 2017

[29] Krugman, Paul. Increasing Returns and Economic Geography. *Journal of Political Economy*. Volume 99: 483 – 499. 1991

[30] Kwak, Sungil and Stephen Smith. Multidimensional poverty and interlocking poverty traps: Framework and application to ethiopian household panel data. Institute for International Economic Policy Working Paper Series, 2011(4). 2011

[31] Kwak, Sungil and Stephen Smith. Regional Agricultural Endowments and Shifts of Poverty Trap Equilibria: Evidence from Ethiopian Panel Data. *The Journal of Development Studies*. 49(7): 955-975. 2013

[32] Laszlo, Sonia. Self-employment earnings and returns to education in rural Peru. *The Journal of Development Studies.* 41(7): 1247-1287. 2005

[33] Lybbert, Travis, Christopher Barrett, Solomon Desta and Layne Coppock. Stochastic wealth dynamics and risk management among a poor population. *The Economic Journal.* 114: 750–777. 2004

[34] Maggio, Guido. Multidimensional analysis of poverty dynamics in Great Britain. ISER working papers. 2004(10). 2004

[35] Matsuyama, Kiminori. Financial Market Globalization, Symmetry-Breaking, and Endogenous Inequality of Nations. *Econometrica.* Volume 72: 853–884. 2004

[36] Maxwell, Daniel, Bapu Vaitla, Girmay Tesfay and Nigussie Abadi. Resilience, Food Security Dynamics, and Poverty Traps in Northern Ethiopia Analysis of a Biannual Panel Dataset, 2011–2013. Working paper, Feinstein International Center. 2013

[37] Mission for the Splicing of Employment, Poverty and Inequality. Monetary Poverty in Colombia: New methodology and 2002-2010 figures. DANE and DNP. 2012

[38] Mookherjee, Dilip and Debraj Ray. Contractual Structure and Wealth Accumulation. *American Economic Review.* 92(4):818-849. 2002

[39] Mukasa, Adamon. Food Price Shocks-induced Poverty Traps: Analysis Using a Panel Dataset From Uganda. Presented in Commodity market instability and asymmetries in developing countries: Development impacts and policies. Clermont-Ferrand, France. 2015

[40] Muñoz, Manuel. Estimación de equivalencias de escala. *Revista de Economía Institucional.* 16(30): 99-130. 2014

[41] Murphy, Kevin, Andrei Shleifer and Robert Vishny. Industrialization and the Big Push. *Journal of Political Economy.* 97(5): 1003 – 1026. 1989

[42] Naschold, Felix. The Poor Stay Poor: Household Asset Poverty Traps in Rural Semi-Arid India. *World Development*, 40(10): 2033-2043. 2012

[43] Pissarides, Christopher and Guglielmo Weber. An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*. 39(1): 17-32. 1989.

[44] Quah, Danny. Convergence Empirics across Economies with (Some) Capital Mobility. *Journal of Economic Growth*.Volume 1: 95 – 124. 1996

[45] Quisumbing, Arnes and Bob Baulch. Assets and Poverty Traps in Rural Bangladesh. *The Journal of Development Studies*. 49(7): 898-916. 2013

[46] Ravallion, Martin. Measuring Social Welfare with and without Poverty Lines. *American Economic Review*. 84(2): 359-364. 1994

[47] Rodriguez, Jorge and Claudio Gonzalez. Household Income Dynamics and Poverty Traps in El Salvador. Presented in AAEA Annual Meetings 2004. The Ohio State University.

[48] Sandoval, Carlos. Household income dynamics and income convergence: A new way to measure poverty. Working Paper. 2014

[49] Universidad de los Andes. Colombia en movimiento: Un analisis descriptivo basado en la Encuesta Longitudinal Colombiana de la Universidad de los Andes ELCA. CEDE. 2011

[50] Yang, Dennis and Mark Yuying. Human capital, entrepreneurship, and farm household earnings. *Journal of Development Economics*. 68(1): 65-88. 2002

[51] You, Jung. Risk, under-investment in agricultural assets and dynamic asset poverty in rural China. *China Economic Review*. 29: 27-45. 2014

[52] Zerfu, Daniel. Poverty Traps and Institutions in Ethiopia. Working paper. African Development Bank. 2012

[53] Zimmerman, Frederick and Carter, Michael. Asset smoothing, consumption smoothing and the reproduction of inequality under risk and subsistence constraints. *Journal of Development Economics*. 71(2): 233–60. 2003

[54] Zhou, Li and Calum Turvey. Testing Asset Dynamics for Poverty Traps in Rural China. *Canadian Journal of Agricultural Economics*. 63(2015): 129–162. 2015

9. Appendix

Appendix 1: FLSS Two step results. 2007-2010. Average Variables

	(1)	(2)	(3)	(4)	(5)
VARIABLES	OLS (2007)	OLS (2010)	FE	IV (2007)	IV (2010)
Age of working members (average)	0.0423	0.1400*	0.0476	-0.0117	0.1112
	(0.052)	(0.073)	(0.037)	(0.059)	(0.075)
Age of working members 2 (average)	-0.0001	-0.0013	-0.0005	0.0009	-0.0008
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Education of working members (average)	0.2306***	0.2860***	-0.0110	0.4980***	0.5119***
	(0.025)	(0.037)	(0.035)	(0.060)	(0.078)
Employment rate among household members	3.4724***	2.8645***	1.6159***	3.5163***	2.6712***
	(0.390)	(0.598)	(0.280)	(0.430)	(0.620)
Proportion of members between 0 and 12 years old	-2.2062***	-1.7788**	-0.4789	-1.0773*	-0.7950
	(0.528)	(0.877)	(0.594)	(0.626)	(0.964)
Proportion of members 62+ years old	-1.4008**	0.6369	0.0950	-2.2026***	0.3908
	(0.597)	(0.764)	(0.559)	(0.677)	(0.791)
Number of self-employed household members	-0.0054	-0.0843	0.2517***	0.0672	-0.1032
	(0.088)	(0.152)	(0.070)	(0.098)	(0.157)
Bogota	-0.2027	0.2270		-0.5495	-0.2357
	(0.421)	(0.651)		(0.469)	(0.686)
Cali	-0.0684	0.2995		-0.4150	-0.1134
	(0.427)	(0.654)		(0.476)	(0.685)
Constant	-3.4745***	-6.1081***	0.1911	-6.2429***	-8.5068***
	(1.211)	(1.795)	(0.997)	(1.445)	(1.997)
Observations	516	516	2,293	511	511
R-squared	0.30	0.18	0.06	0.14	0.12
F	23.80	12.24	9.938	.	.
Number of id_hogar1			637		
Year FE			YES		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ The table presents the first step in the multi-step approach proposed by Adato, Carter and May (2006). Columns 1 and 2 present the results using OLS for estimating the first step for 2007 and 2010 separately. Column 3 shows the results, estimating the first step model using all the years available (2007, 2008, 2009, 2010) through an FE data panel model. Columns 4 and 5 estimate the first step model using IV for the years 2007 and 2010 assuming that the education of the head of the household shows reverse causality with respect to the adult equivalent income divided by the poverty line. Education is instrumented with variables of the parents of the head of the household such as education of the head of the household where the head of the household grew up, the working status of the head of the household where the head of the household grew up and the self-reported economic status of the household where the head of the household grew up.

Appendix 2: FLSS Two step results. 2007-2010. Total Variables

	(1)	(2)	(3)	(4)	(5)
VARIABLES	OLS (2007)	OLS (2010)	FE	IV (2007)	IV (2010)
Age of working members (total)	-0.0107*	-0.0316***	0.0042	-0.0662***	-0.0972***
	(0.006)	(0.010)	(0.004)	(0.013)	(0.018)
Age of working members 2 (total)	-0.0000	0.0001**	-0.0000	-0.0000	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education of working members (total)	0.0602***	0.0831***	-0.0019	0.3040***	0.4030***
	(0.010)	(0.016)	(0.009)	(0.043)	(0.069)
Employment rate among household members	3.2553***	2.0321***	1.5159***	2.8550***	1.1441
	(0.362)	(0.549)	(0.253)	(0.500)	(0.711)
Proportion of members between 0 and 12 years old	-2.6171***	-2.6538***	-0.2232	-1.3912**	-0.4688
	(0.466)	(0.772)	(0.551)	(0.677)	(1.075)
Proportion of members 62+ years old	-0.2512	0.6085	-0.0394	1.2203**	2.2822***
	(0.321)	(0.442)	(0.423)	(0.507)	(0.652)
Number of self-employed household members	-0.1176	-0.3623**	0.2885***	-0.7566***	-1.6209***
	(0.094)	(0.159)	(0.070)	(0.167)	(0.327)
Bogota	-0.0300	0.3261		-0.5261	-0.9491
	(0.388)	(0.588)		(0.537)	(0.781)
Cali	0.0999	0.4515		-0.2795	-0.6547
	(0.390)	(0.589)		(0.537)	(0.772)
Constant	0.5327	1.3890**	0.7326***	0.1544	1.6110**
	(0.431)	(0.656)	(0.193)	(0.593)	(0.821)
Observations	684	684	2,739	677	677
R-squared	0.25	0.11	0.09		
F	24.92	9.636	19.65	.	.
Number of id_hogar1			685		
Year FE			YES		

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ The table presents the first step in the multi-step approach proposed by Adato, Carter and May (2006). Columns 1 and 2 present the results using OLS for estimating the first step for 2007 and 2010 separately. Column 3 shows the results, estimating the first step model using all the years available (2007, 2008, 2009, 2010) through an FE data panel model. Columns 4 and 5 estimate the first step model using IV for the years 2007 and 2010 assuming that the education of the head of the household shows reverse causality with respect to the adult equivalent income divided by the poverty line. Education is instrumented with variables of the parents of the head of the household such as education of the head of the household where the head of the household grew up, the working status of the head of the household where the head of the household grew up and the self-reported economic status of the household where the head of the household grew up.

Appendix 3: FLSS Semi-parametric results. 2007-2010. Different Aggregations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
VARIABLES	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.
Age of working members (average)	0.1287***			0.1134**			0.1095**			0.1134**			0.1095**		
	(0.047)			(0.045)			(0.044)			(0.045)			(0.044)		
Age of working members 2 (average)	-0.0014**			-0.0012**			-0.0012**			-0.0012**			-0.0012**		
	(0.001)			(0.001)			(0.001)			(0.001)			(0.001)		
Education of working members (average)	0.1686***			0.1575***			0.1580***			0.1575***			0.1580***		
	(0.041)			(0.041)			(0.042)			(0.041)			(0.042)		
Employment rate among household members	1.5184**	0.7494*	1.3075***	1.5404**	0.7458*	1.2668***	1.4640**	0.7131*	1.2336***	1.5404**	0.7458*	1.2668***	1.4640**	0.7131*	1.2336***
	(0.669)	(0.451)	(0.494)	(0.630)	(0.422)	(0.442)	(0.616)	(0.425)	(0.437)	(0.630)	(0.422)	(0.442)	(0.616)	(0.425)	(0.437)
Proportion of members between 0 and 12 years old	-1.4083**	-1.6220***	-1.4766***	-1.4903**	-1.6008***	-1.4546***	-1.4197**	-1.5622***	-1.4077***	-1.4903**	-1.6008***	-1.4546***	-1.4197**	-1.5622***	-1.4077***
	(0.674)	(0.473)	(0.488)	(0.640)	(0.443)	(0.455)	(0.640)	(0.452)	(0.461)	(0.640)	(0.443)	(0.455)	(0.640)	(0.452)	(0.461)
Proportion of members 62+ years old	1.1228*	0.4376	-0.0478	0.8655	0.3721	-0.0696	0.8656	0.3877	-0.0300	0.8655	0.3721	-0.0696	0.8656	0.3877	-0.0300
	(0.606)	(0.451)	(0.476)	(0.551)	(0.435)	(0.449)	(0.558)	(0.436)	(0.447)	(0.551)	(0.435)	(0.449)	(0.558)	(0.436)	(0.447)
Number of self-employed household members	0.1086	-0.0874	0.2200	0.1050	-0.0972	0.1986	0.1071	-0.0938	0.1990	0.1050	-0.0972	0.1986	0.1071	-0.0938	0.1990
	(0.155)	(0.222)	(0.142)	(0.149)	(0.219)	(0.136)	(0.148)	(0.216)	(0.135)	(0.149)	(0.219)	(0.136)	(0.148)	(0.216)	(0.135)
Bogota	0.1562	0.3211	0.4171	0.2122	0.3041	0.3903	0.1262	0.2777	0.3575	0.2122	0.3041	0.3903	0.1262	0.2777	0.3575
	(0.375)	(0.324)	(0.312)	(0.352)	(0.318)	(0.309)	(0.327)	(0.298)	(0.287)	(0.352)	(0.318)	(0.309)	(0.327)	(0.298)	(0.287)
Cali	0.0347	0.2638	0.2666	0.0506	0.2106	0.2069	-0.0408	0.1781	0.1689	0.0506	0.2106	0.2069	-0.0408	0.1781	0.1689
	(0.334)	(0.314)	(0.281)	(0.310)	(0.306)	(0.273)	(0.294)	(0.285)	(0.247)	(0.310)	(0.306)	(0.273)	(0.294)	(0.285)	(0.247)
Age of working members (total)		-0.0190			-0.0206			-0.0199			-0.0206			-0.0199	
		(0.012)			(0.013)			(0.012)			(0.013)			(0.012)	
Age of working members 2 (total)		0.0001			0.0001			0.0001			0.0001			0.0001	
		(0.000)			(0.000)			(0.000)			(0.000)			(0.000)	
Education of working members (total)		0.0515***			0.0477***			0.0470***			0.0477***			0.0470***	
		(0.016)			(0.016)			(0.017)			(0.016)			(0.017)	
Age of head of the household			0.0347			0.0380			0.0395			0.0380			0.0395
			(0.029)			(0.031)			(0.030)			(0.031)			(0.030)
Age of head of the household 2			-0.0000			-0.0001			-0.0001			-0.0001			-0.0001
			(0.000)			(0.000)			(0.000)			(0.000)			(0.000)
Education of head of the household			0.1412***			0.1372***			0.1377***			0.1372***			0.1377***
			(0.027)			(0.027)			(0.027)			(0.027)			(0.027)
Observations	553	684	684	553	684	684	553	684	684	553	684	684	553	684	684
R-squared	0.09	0.08	0.09	0.08	0.08	0.09	0.08	0.07	0.09	0.08	0.08	0.09	0.08	0.07	0.09
Method	Epan.	Epan.	Epan.	cosine .	cosine .	cosine .	gaussian .	gaussian .	gaussian .	cosine .	cosine .	cosine .	gaussian .	gaussian .	gaussian .
Degree	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
F	9.509	8.049	11.36	9.200	7.626	11.01	8.845	7.358	10.80	9.200	7.626	11.01	8.845	7.358	10.80

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ The dependent variable is the predicted adult equivalent per capita income in 2010, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2007. 3/ It also includes as controls, in other specifications, total and average variables of the household.

Appendix 4: FLSS Underreport Income Correction Regressions . 2007-2010

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS (Def 2-Unrest. 25%)	OLS (Def 2-Rest. 25%)	BE (Def 1-Unrest.)	IV (Def 1-Unrest.)	IV (Def 1-Rest.)	IV (Def 2-Unrest. 100%)
Log total household income			0.1117***	0.3999***	0.3142***	0.3965***
			(0.021)	(0.049)	(0.060)	(0.049)
Self-employed household (Def.1 - Unrest.)			0.1276**	0.0802**		
			(0.061)	(0.039)		
Bogota	0.2917**	0.5319***	-0.0124	0.0355	0.1392	0.0364
	(0.118)	(0.187)	(0.118)	(0.093)	(0.114)	(0.093)
Cali	0.3587***	0.6672***	0.0510	0.0274	0.1035	0.0321
	(0.120)	(0.191)	(0.120)	(0.096)	(0.120)	(0.096)
2008	-0.0621	-0.0677	-0.5795***	-0.0692	-0.0438	-0.0730
	(0.046)	(0.078)	(0.170)	(0.053)	(0.073)	(0.052)
2009			-0.0887	0.0050	-0.0008	0.0015
			(0.155)	(0.053)	(0.073)	(0.053)
2010			0.1999	0.1129**	0.1340*	0.1092**
			(0.153)	(0.053)	(0.074)	(0.053)
Age of head of the household	0.0011	0.0167	-0.0127	-0.0110	0.0094	-0.0103
	(0.014)	(0.025)	(0.014)	(0.011)	(0.017)	(0.011)
Age of head of the household 2	0.0001	-0.0001	0.0002	0.0002*	0.0000	0.0002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married head of the household	0.1711***	0.2568***	0.1741***	0.0387	0.0528	0.0473
	(0.050)	(0.085)	(0.054)	(0.045)	(0.057)	(0.045)
Household size	0.0498***	0.0662***	0.0541***	0.0034	0.0005	0.0072
	(0.014)	(0.023)	(0.014)	(0.012)	(0.016)	(0.012)
Hours worked weekly (Head of the Household)	-0.0001	-0.0008	0.0018	0.0008	0.0017	0.0008
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Education of head of the household	0.0241***	0.0241**	0.0294***			
	(0.006)	(0.011)	(0.006)			
Log total household income (3 years average)	0.1936***	0.1317***				
	(0.030)	(0.047)				
Self-employed household (Def.2 - Unrest. 25%)	0.0938*					
	(0.048)					
Self-employed household (Def.2 - Rest. 25%)		0.1589*				
		(0.086)				
Self-employed household (Def.1 - Rest.)					0.1193**	
					(0.056)	
Self-employed household (Def.2 - Rest. 100%)						0.0753*
						(0.044)
Constant	8.9398***	9.1089***	10.5719***	7.2175***	7.7103***	7.2471***
	(0.501)	(0.824)	(0.429)	(0.629)	(0.853)	(0.631)
Observations	868	445	1,731	1,712	867	1,712
R-squared	0.18	0.14	0.24			
Underreport Rate %	38.4	70.09	68.11	18.17	31.6	17.3
F	16.56	6.180	12.75	.	.	.
Number of id_hogar1			535			

Standard errors in parentheses

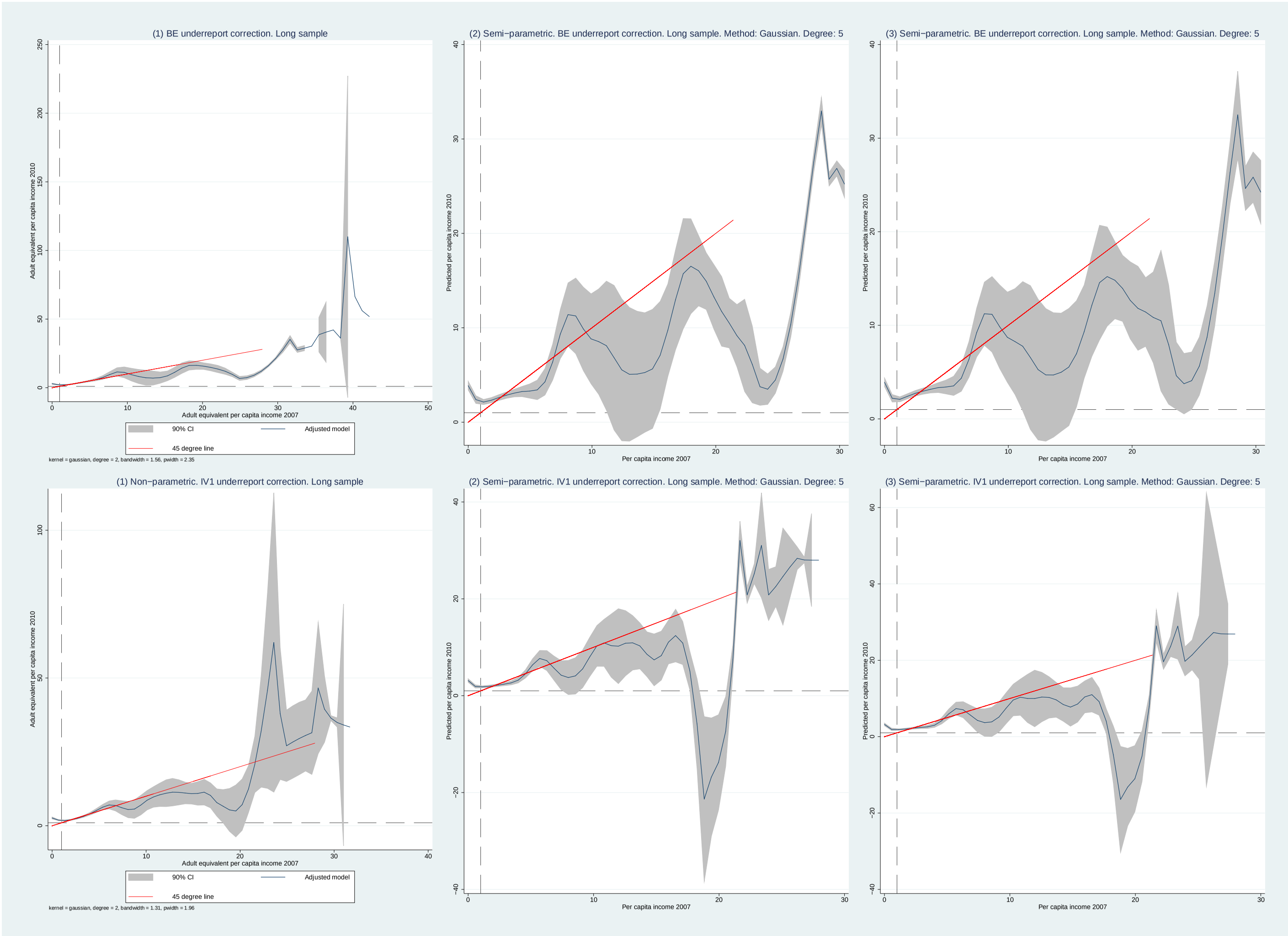
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Results of the total income underreport estimation using different estimation methods (OLS, BE and IV) and different self-employed household definitions. 2/ The self-employment definitions include many variations. Definition 1 implies that a household is self-employed if at least one of the adult members of the household report being self-employed. In this case a person is considered to be self-employed if he works as a pawn, he works on his own business, or he is an employer. In definition 2, I use the proportion of the income that is derived from business and the fees received from independent work as a measure of the degree on how self-employed the household is. If said proportion is higher than 25% the household is classified as self-employed. 3/ Following Kukk and Staehr (2014) a household may be considered self-employed only when it is self-employed in every single period. In the same way, a household is not self-employed when it is not self-employed in any of the observed periods. This definition implies that the households changing self-employed status through time should be ignored, restricting the sample size. In contraposition to this, if the household is considered self-employed period by period individually, the sample is unrestricted.

Figure 12: Appendix 4A

Non parametric and semi parametric regressions of the Household per Capita Income Divided by the Poverty Line Correcting for Income Underreport

2007 (Horizontal axis) vs. 2010 (Vertical axis).FLSS: Unrestricted sample

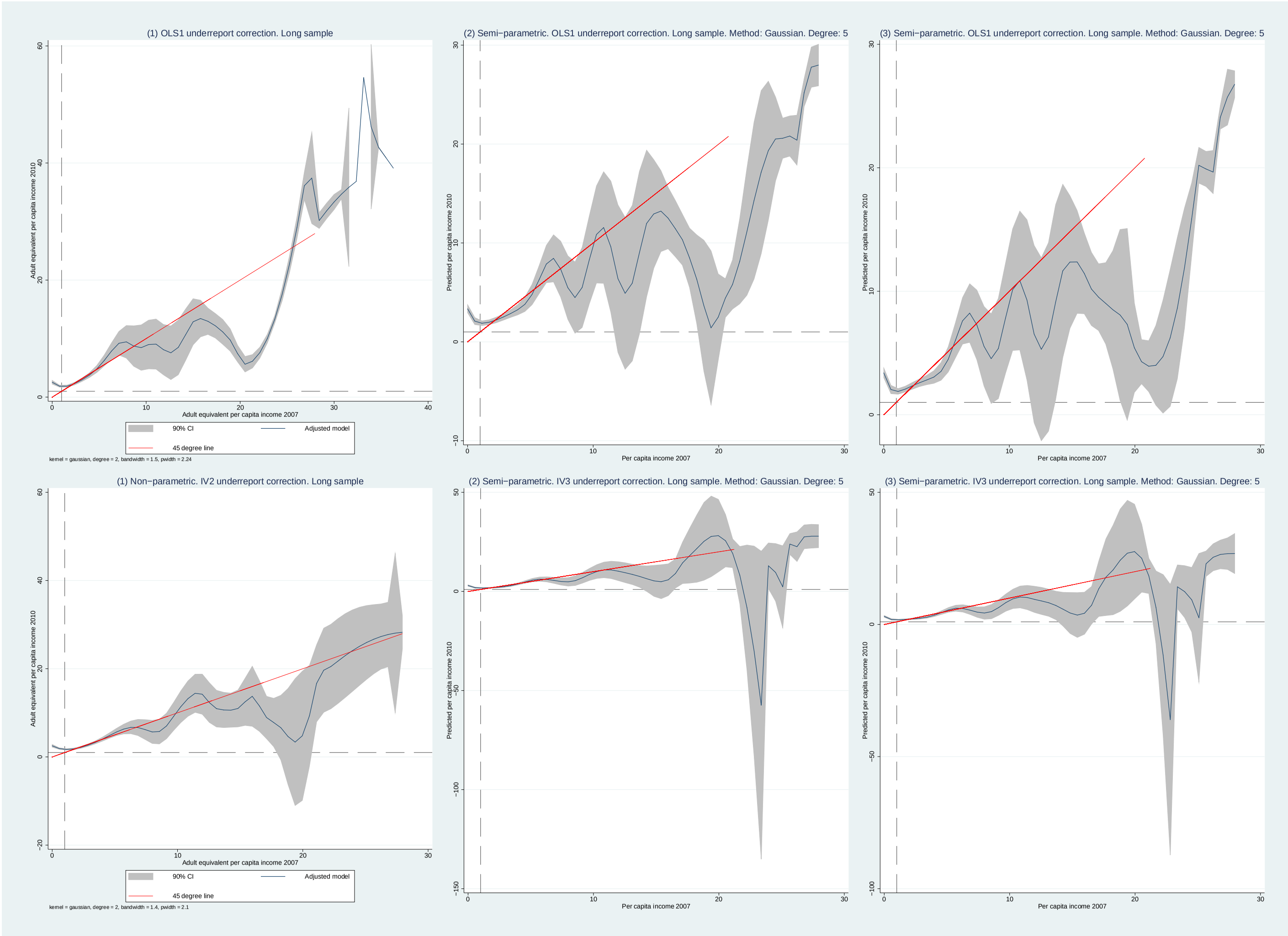


Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS). Note: 1/ For non-parametric regressions: Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2010, on the same variable for the year 2007. 2/ For semi-parametric regressions:Result using the Robinson's semiparametric estimator. Variable in the vertical axis is the predicted adult equivalent per capita income in 2010, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2007. 3/ In both types of model a rule of thumb were used to select the bandwidth and a fifth-degree polynomial was used in the smoothing process. 4/Both models use a Gaussian kernel. 5/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Figure 13: Appendix 4B

Non parametric and semi parametric regressions of the Household per Capita Income Divided by the Poverty Line Correcting for Income Underreport

2007 (Horizontal axis) vs. 2010 (Vertical axis).FLSS: Unrestricted sample

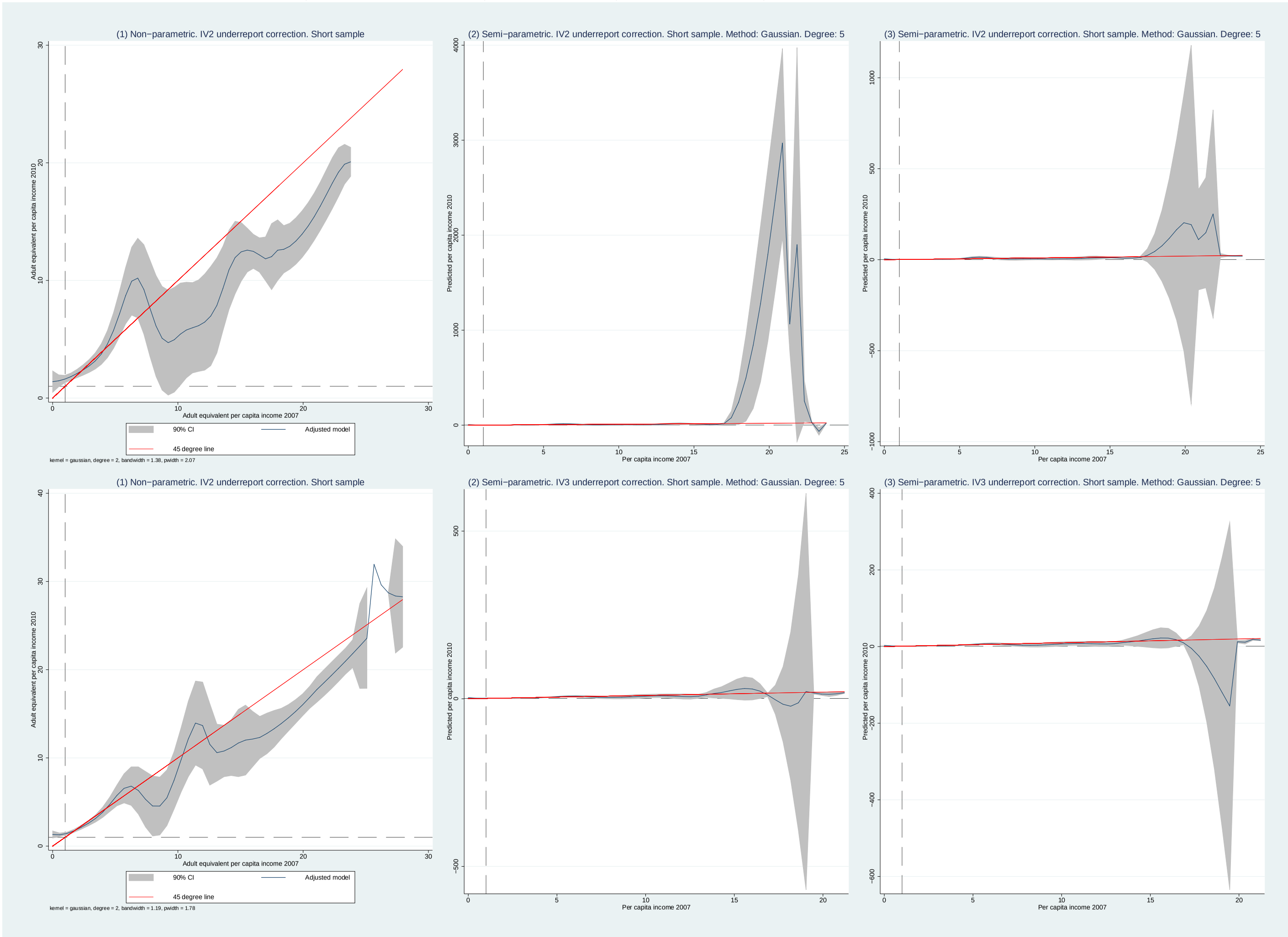


Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS). Note: 1/ For non-parametric regressions: Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2010, on the same variable for the year 2007. 2/ For semi-parametric regressions:Result using the Robinson's semiparametric estimator. Variable in the vertical axis is the predicted adult equivalent per capita income in 2010, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2007. 3/ In both types of model a rule of thumb were used to select the bandwidth and a fifth-degree polynomial was used in the smoothing process. 4/Both models use a Gaussian kernel. 5/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Figure 14: Appendix 4C

Non parametric and semi parametric regressions of the Household per Capita Income Divided by the Poverty Line Correcting for Income Underreport

2007 (Horizontal axis) vs. 2010 (Vertical axis).FLSS: Restricted sample

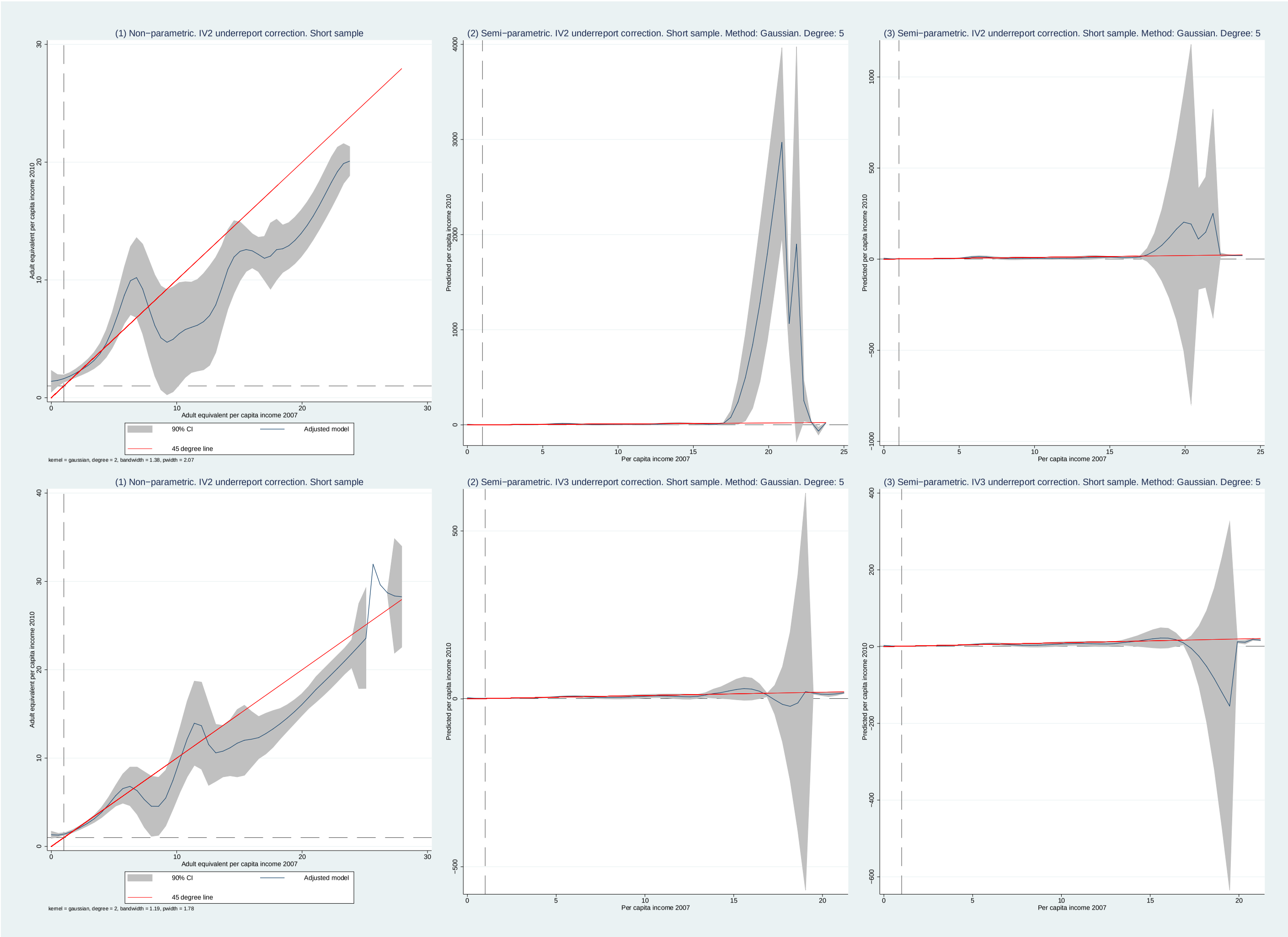


Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS). Note: 1/ For non-parametric regressions: Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2010, on the same variable for the year 2007. 2/ For semi-parametric regressions:Result using the Robinson's semiparametric estimator. Variable in the vertical axis is the predicted adult equivalent per capita income in 2010, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2007. 3/ In both types of model a rule of thumb were used to select the bandwidth and a fifth-degree polynomial was used in the smoothing process. 4/Both models use a Gaussian kernel. 5/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Figure 15: Appendix 4D

Non parametric and semi parametric regressions of the Household per Capita Income Divided by the Poverty Line Correcting for Income Underreport

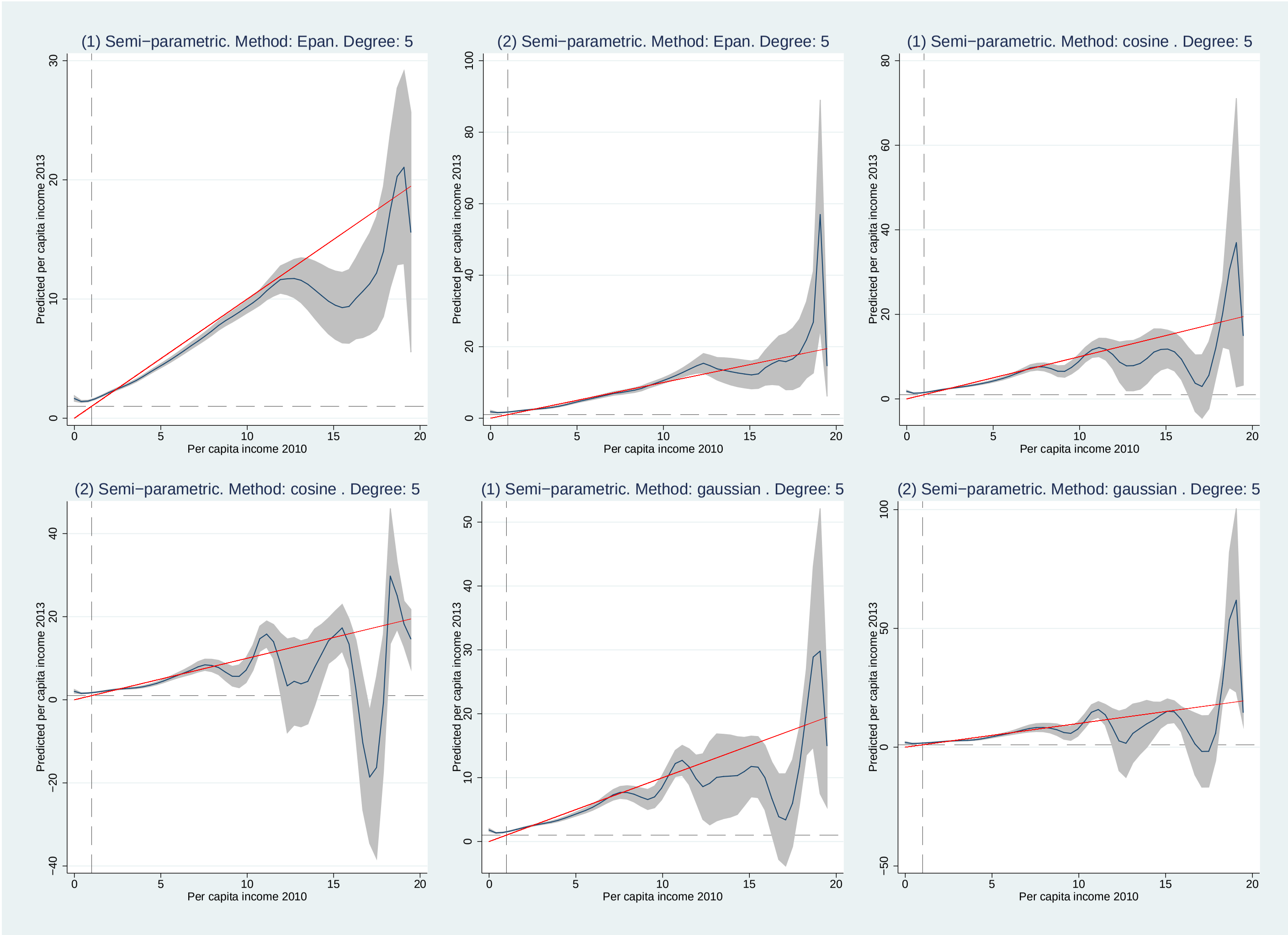
2007 (Horizontal axis) vs. 2010 (Vertical axis).FLSS: Restricted sample



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS). Note: 1/ For non-parametric regressions: Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2010, on the same variable for the year 2007. 2/ For semi-parametric regressions:Result using the Robinson's semiparametric estimator. Variable in the vertical axis is the predicted adult equivalent per capita income in 2010, while the controls includes in the parametric part the same controls included in Table 2 and in the non-parametric part the observed adult equivalent per capita income in 2007. 3/ In both types of model a rule of thumb were used to select the bandwidth and a fifth-degree polynomial was used in the smoothing process. 4/Both models use a Gaussian kernel. 5/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Figure 16: Appendix 5A

Vertical axis: Predicted adult equivalent per capita income divided by the poverty line in 2013
CLSA: Semi-parametric results. 2010-2013.All Urban Areas



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ Endogenous variable is the predicted adult equivalent per capita income in 2013, while the controls includes, in the parametric part, the same controls included in Table 3, and, in the non-parametric part, the observed adult equivalent per capita income in 2010. 3/ All graphs use a rule of thumb for choosing the bandwidth and a polynomial approximation of the 5th degree. The graphs use cosine, Epanechnikov and gaussian kernel. 4/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Appendix 5B: CLSA Semi-parametric results. 2010-2013. All Urban Areas

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.
Age of head of the household	0.0018	0.0080	-0.0003	0.0060	-0.0020	0.0060
	(0.019)	(0.032)	(0.019)	(0.030)	(0.019)	(0.030)
Age of head of the household 2	0.0002	0.0000	0.0002	0.0000	0.0002	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education of head of the household	0.0872***	0.0491**	0.0835***	0.0502**	0.0831***	0.0501**
	(0.010)	(0.020)	(0.010)	(0.020)	(0.010)	(0.020)
Self-employed head of the household	0.0255	0.0814	0.0398	0.0858	0.0357	0.0820
	(0.072)	(0.108)	(0.071)	(0.104)	(0.071)	(0.103)
Age of head of household's partner		-0.0252		-0.0181		-0.0200
		(0.031)		(0.030)		(0.030)
Age of head of household's partner 2		0.0004		0.0003		0.0003
		(0.000)		(0.000)		(0.000)
Education of head of household's partner		0.0671***		0.0630***		0.0645***
		(0.015)		(0.015)		(0.015)
Self-employed head of household's partner		0.0110		-0.0226		-0.0050
		(0.112)		(0.106)		(0.108)
Proportion of members between 0 and 12 years old	-1.1907***	-1.5293***	-1.2490***	-1.5604***	-1.2444***	-1.5660***
	(0.193)	(0.351)	(0.190)	(0.326)	(0.189)	(0.328)
Proportion of members 62+ years old	-0.1034	-0.2481	-0.1688	-0.2229	-0.1737	-0.2965
	(0.281)	(0.527)	(0.275)	(0.497)	(0.274)	(0.491)
Observations	3,326	1,402	3,326	1,402	3,326	1,402
R-squared	0.09	0.12	0.09	0.13	0.09	0.12
City FE	YES	YES	YES	YES	YES	YES
Method	Epan.	Epan.	cosine	cosine	gaussian	gaussian
Degree	5	5	5	5	5	5
F	2946	79.54	2458	41.84	4510	55.51

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

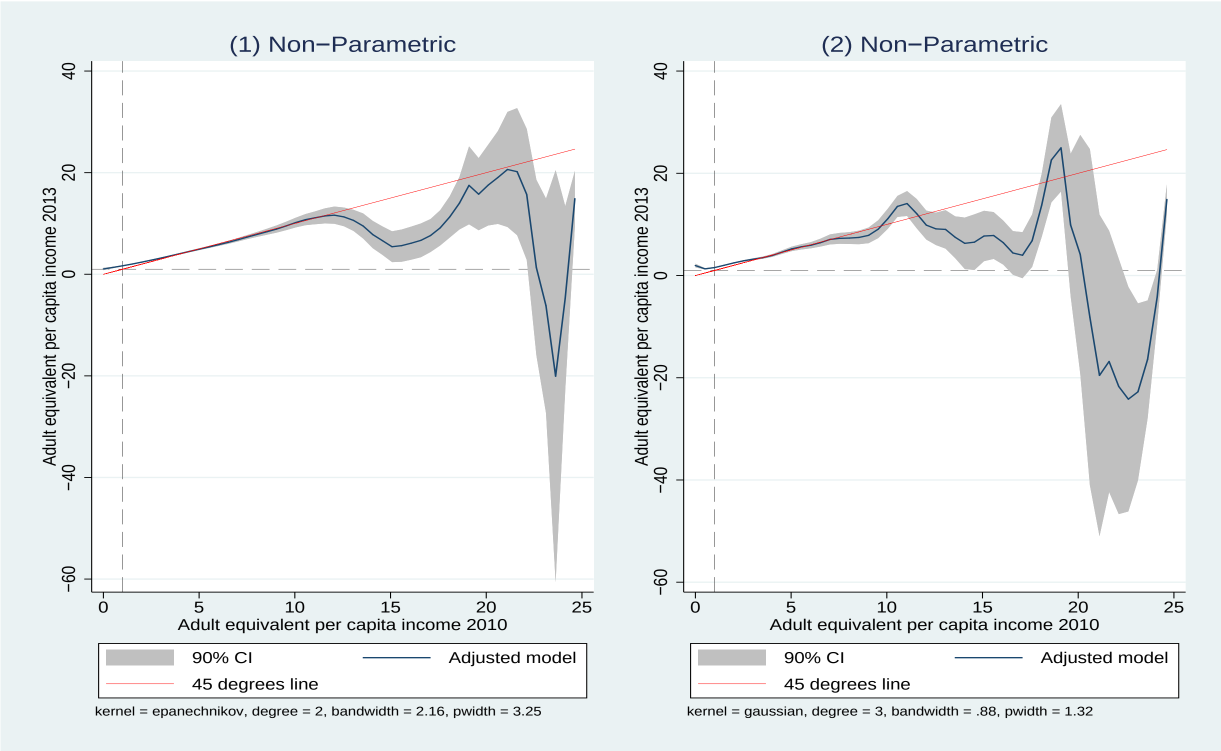
Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ The dependent variable is the adult equivalent per capita income in 2013, while the controls includes in the parametric part the same controls included in Table 3 and in the non-parametric part the observed adult equivalent per capita income in 2010. 3/ It also includes as controls, in other specifications, partner of the head of the household variables. 4/ All urban areas

Figure 17: Appendix 6A

Non parametric regression of the Household per Capita Income Divided by the Poverty Line

2010 (Horizontal axis) vs. 2013 (Vertical axis)

CLSA: Colombian Longitudinal Survey by Universidad los Andes. 14 main cities. 2010-2013



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes (CLSA). Note: 1/ Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2013, on the same variable for the year 2010. 2/ In both graphs a rule of thumb were used to select the bandwidth and a second-degree polynomial was used in the smoothing process. 3/ The left graph was estimated using an Epanechnikov kernel while the graph on the right use a Gaussian kernel. 4/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Appendix 6B: CLSA Two step results. 2010-2013. 14 main cities

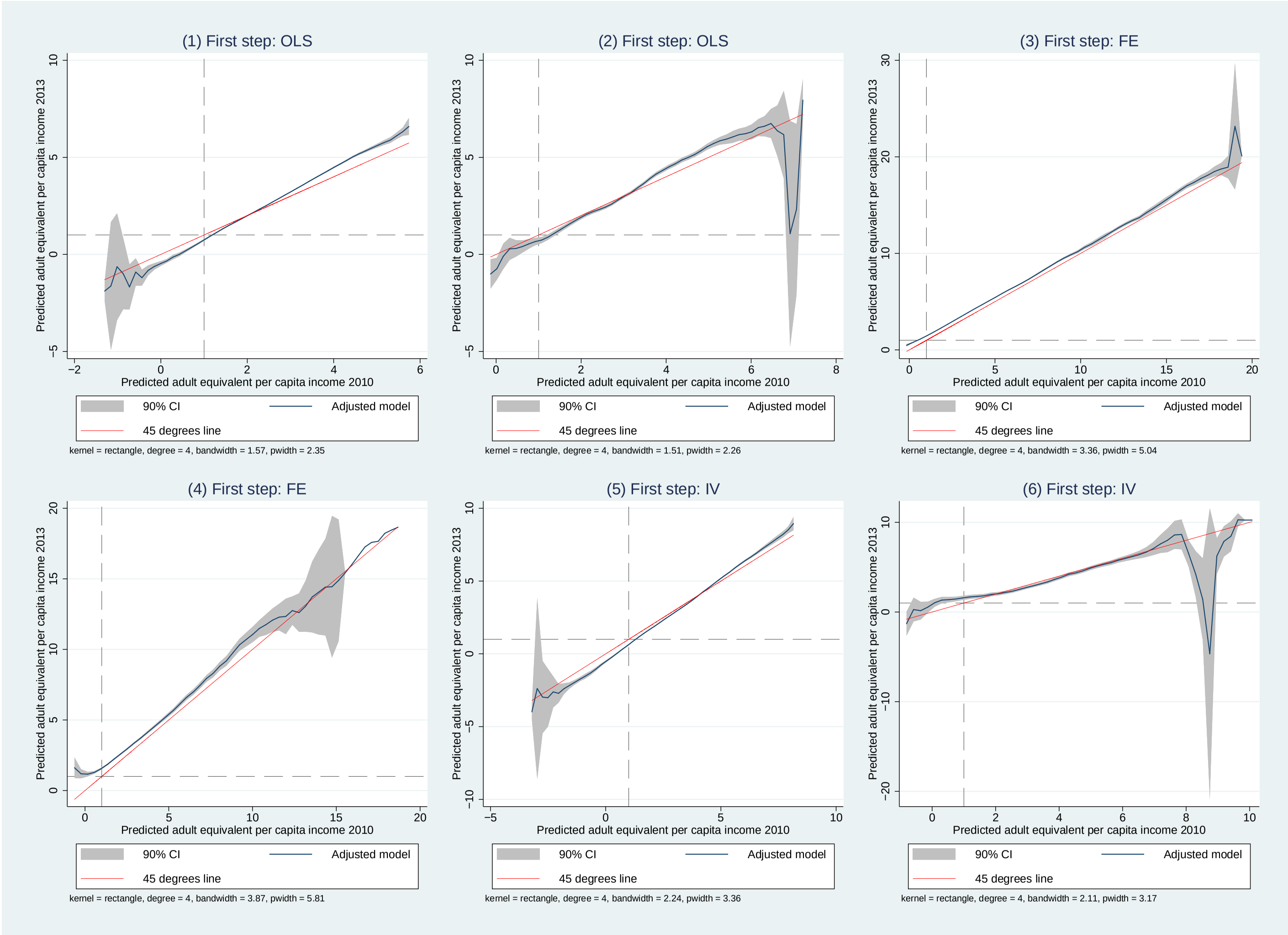
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	OLS (2010)	OLS (2013)	OLS (2010)	OLS (2013)	FE	FE	IV (2010)	IV (2013)	IV (2010)	IV (2013)
Age of head of the household	0.0144	-0.0413	0.1197	0.2409	-0.0726	-0.2415	0.0212	-0.0718	0.0913	0.1532
	(0.033)	(0.055)	(0.104)	(0.199)	(0.062)	(0.199)	(0.046)	(0.079)	(0.137)	(0.277)
Age of head of the household 2	0.0000	0.0008	-0.0012	-0.0025	0.0004	0.0021	0.0000	0.0012	-0.0005	-0.0015
	(0.000)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)
Education of head of the household	0.1997***	0.2472***	0.0962***	0.1330***	-0.0044	-0.1923**	0.3601***	0.4079***	0.1931***	0.2039*
	(0.012)	(0.016)	(0.031)	(0.049)	(0.036)	(0.085)	(0.032)	(0.044)	(0.074)	(0.116)
Self-employed head of the household	0.0725	0.2103	-0.0980	0.3526	0.0294	0.0439	0.3175**	0.4065**	-0.0888	0.2560
	(0.104)	(0.145)	(0.209)	(0.336)	(0.119)	(0.255)	(0.142)	(0.196)	(0.272)	(0.436)
Age of head of household's partner			-0.1187	-0.1490		0.0749			-0.1127	-0.0286
			(0.095)	(0.167)		(0.176)			(0.136)	(0.224)
Age of head of household's partner 2			0.0015	0.0019		0.0002			0.0015	0.0007
			(0.001)	(0.002)		(0.002)			(0.002)	(0.002)
Education of head of household's partner			0.1933***	0.2410***		0.0884			0.2217***	0.1229
			(0.031)	(0.052)		(0.073)			(0.066)	(0.120)
Self-employed head of household's partner			0.0570	0.1435		0.0297			-0.0138	-0.2265
			(0.210)	(0.334)		(0.244)			(0.291)	(0.471)
Proportion of members between 0 and 12 years old	-1.3778***	-2.0054***	-0.9753	-2.0168*	-0.7752*	-1.2471	-1.4616***	-2.1817***	-0.7841	-3.0604**
	(0.284)	(0.417)	(0.634)	(1.047)	(0.397)	(0.929)	(0.378)	(0.532)	(0.867)	(1.388)
Proportion of members 62+ years old	0.0778	0.2098	-1.5521	-1.1553	0.4155	-1.6061	1.0989	0.7935	-4.0355	-2.6353
	(0.417)	(0.538)	(1.312)	(1.678)	(0.523)	(1.261)	(0.825)	(0.852)	(3.646)	(2.799)
Constant	-0.7172	-0.4175	-1.5297	-4.9510	3.6621*	7.1806	-2.9534***	-2.0990	-3.7360	-5.0996
	(0.718)	(1.285)	(1.821)	(3.680)	(2.019)	(5.404)	(1.033)	(1.832)	(2.667)	(5.167)
Observations	1,373	1,373	351	351	2,746	702	942	942	207	207
R-squared	0.23	0.21	0.31	0.23	0.05	0.09	0.19	0.15	0.36	0.25
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
F	21.04	18.60	6.499	4.188	3.548	2.744	.	.	.	.
Number of consecutivo					1,373	351				
Year FE					YES	YES				

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ The table presents the first step in the multi-step approach proposed by Adato, Carter and May (2006). Columns 1 and 2 present the results using OLS for estimating the first step for 2010 and 2013 separately. Column 5 show the results, estimating the first step model using all the years available as a pool (2010 and 2013) through an FE data panel model. Columns 7 and 8 estimates the first step model using IV for the years 2010 and 2013 assuming that the education of the head of the household shows reversal causality in respect to the adult equivalent income divided by the poverty line. Education is instrumented with variables of the parents of the head of the household. The other columns includes as additional controls variables of the spouse of the head of the household. 2/ 14 main cities

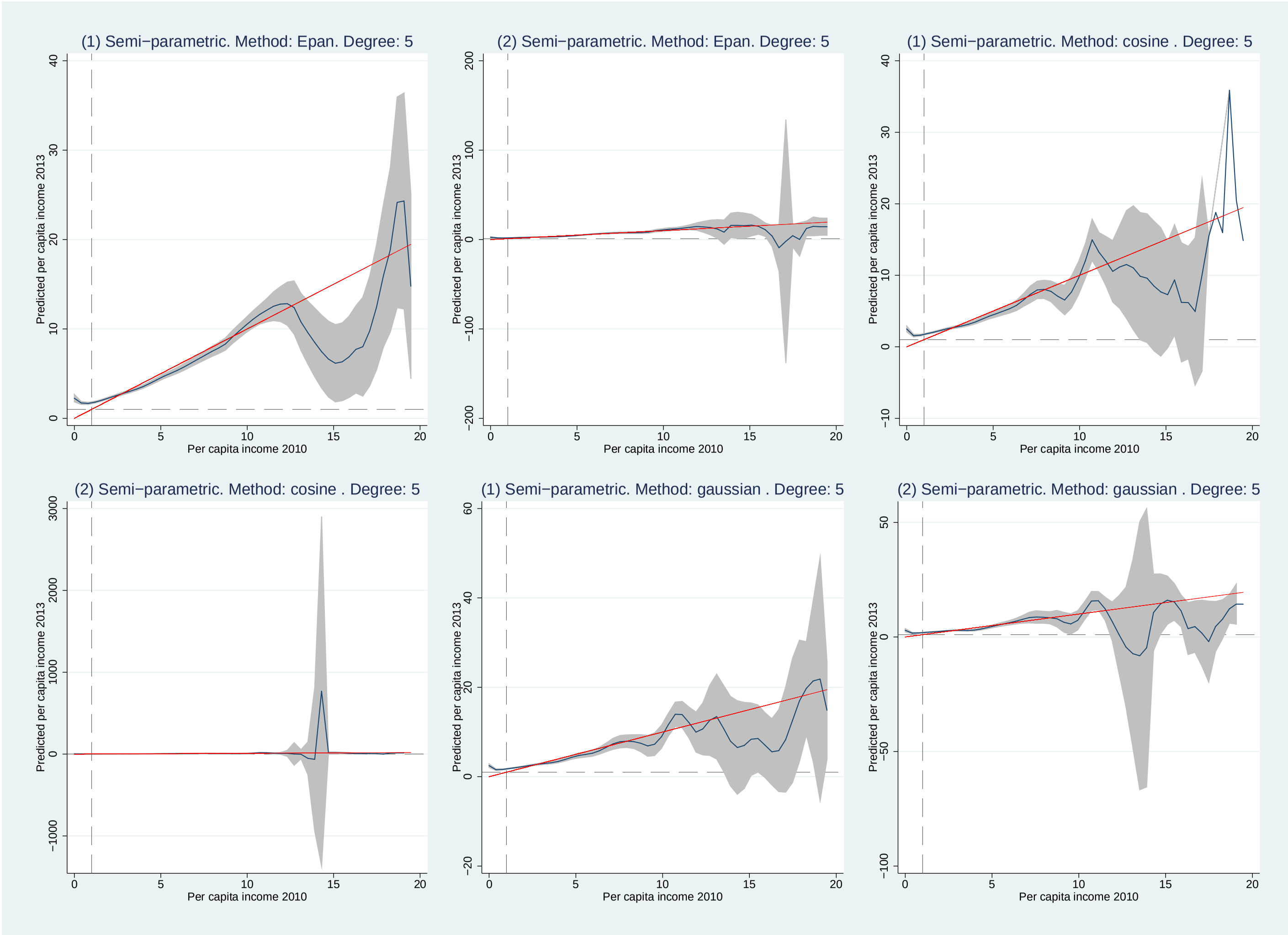
Figure 18: Appendix 6C: Colombian Longitudinal Survey by Universidad los Andes. Second step of the two step method. 2010-2013. 14 main cities



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ The graph presents the second step in the multi-step approach proposed by Adato, Carter and May (2006). The vertical axis variable is the predicted adult equivalent per capita income divided by the poverty line in 2013 and the horizontal axis variable is the prediction for the same variable in 2010. 2/ The non parametric model was estimated using a kernel-weighted local polynomial smoothing regression with a rectangle kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb. 3/ Graphs 1 and 2 use the OLS first step results, while graphs 3 and 4 uses the FE first step results and graphs 5 and 6 uses the first step IV results. Odd graphs use only head of the household controls, while even graphs use head of the household and its spouse controls. 4/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line. 5/ 14 main cities

Figure 19: Appendix 6D

Vertical axis: Predicted adult equivalent per capita income divided by the poverty line in 2013
CLSA: Semi-parametric results. 2010-2013. 14 main cities



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ Endogenous variable is the predicted adult equivalent per capita income in 2013, while the controls include, in the parametric part, the same controls included in Table 3, and, in the non-parametric part, the observed adult equivalent per capita income in 2010. 3/ All graphs use a rule of thumb for choosing the bandwidth and a polynomial approximation of the 5th degree. The graphs use cosine, Epanechnikov and gaussian kernel. 4/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Appendix 6E: CLSA Semi-parametric results. 2010-2013. 14 main cities

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.
Age of head of the household	-0.0314	-0.0115	-0.0295	0.0021	-0.0299	-0.0013
	(0.035)	(0.059)	(0.035)	(0.058)	(0.035)	(0.058)
Age of head of the household 2	0.0005	0.0001	0.0004	-0.0000	0.0004	0.0000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Education of head of the household	0.1044***	0.0545	0.1013***	0.0547	0.1014***	0.0552
	(0.017)	(0.036)	(0.016)	(0.036)	(0.016)	(0.036)
Self-employed head of the household	0.1331	0.1370	0.1273	0.1447	0.1287	0.1490
	(0.101)	(0.139)	(0.099)	(0.140)	(0.099)	(0.139)
Age of head of household's partner		-0.0285		-0.0291		-0.0272
		(0.052)		(0.049)		(0.050)
Age of head of household's partner 2		0.0005		0.0005		0.0005
		(0.001)		(0.001)		(0.001)
Education of head of household's partner		0.0852***		0.0833***		0.0842***
		(0.024)		(0.024)		(0.024)
Self-employed head of household's partner		0.0505		0.0143		0.0268
		(0.158)		(0.150)		(0.150)
Proportion of members between 0 and 12 years old	-1.6112***	-1.8055***	-1.6291***	-1.9554***	-1.6380***	-1.9600***
	(0.302)	(0.510)	(0.296)	(0.473)	(0.295)	(0.473)
Proportion of members 62+ years old	-0.0260	-0.7898	-0.0531	-0.7256	-0.0461	-0.7444
	(0.435)	(0.602)	(0.434)	(0.614)	(0.433)	(0.607)
Observations	1,720	734	1,720	733	1,720	734
R-squared	0.07	0.08	0.07	0.08	0.07	0.09
City FE	YES	YES	YES	YES	YES	YES
Method	Epan.	Epan.	cosine	cosine	gaussian	gaussian
Degree	5	5	5	5	5	5
F	7.264	2.857	7.016	3.041	7.011	3.019

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

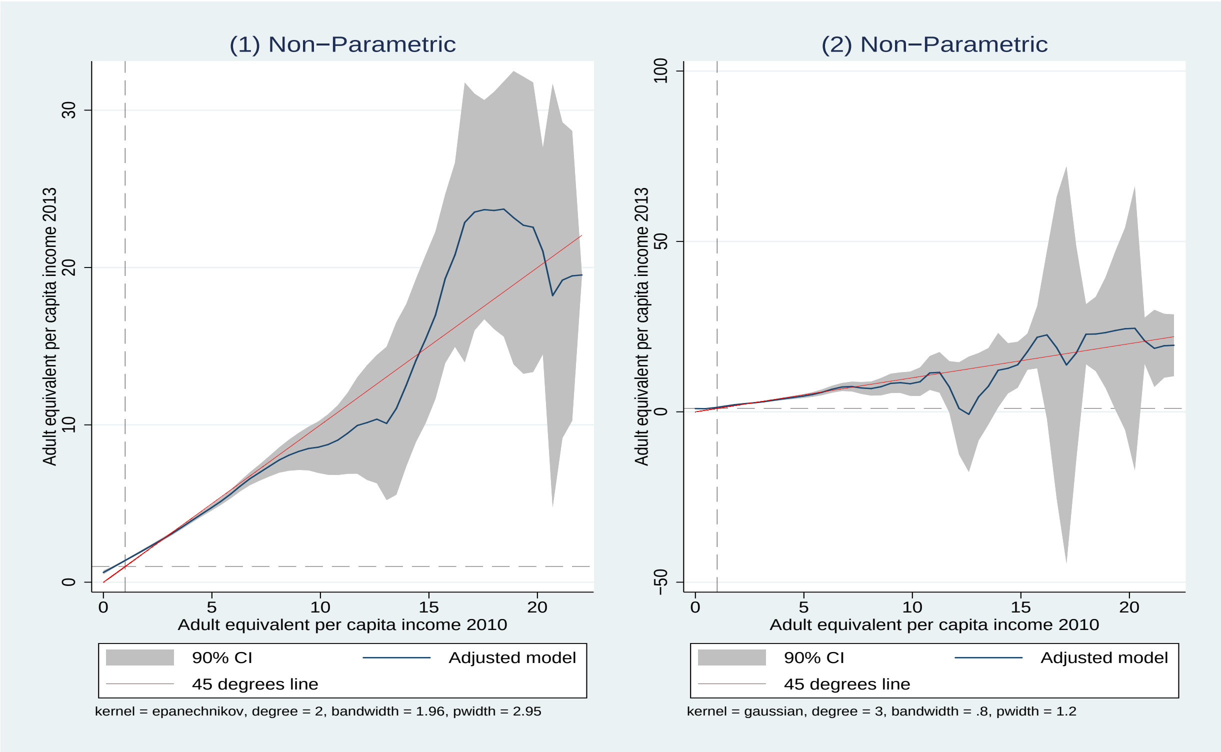
Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ The dependent variable is the adult equivalent per capita income in 2013, while the controls includes in the parametric part the same controls included in Table 3 and in the non-parametric part the observed adult equivalent per capita income in 2010. 3/ It also includes as controls, in other specifications, partner of the head of the household variables. 4/ 14 main cities

Figure 20: Appendix 7A: CLSA

Non parametric regression of the Household per Capita Income Divided by the Poverty Line

2010 (Horizontal axis) vs. 2013 (Vertical axis)

CLSA: Colombian Longitudinal Survey by Universidad los Andes. Other urban areas. 2010-2013



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes (CLSA). Note: 1/ Kernel-weighted local polynomial smoothing regression of the adult equivalent per capita income per household divided by the poverty line in 2013, on the same variable for the year 2010. 2/ In both graphs a rule of thumb were used to select the bandwidth and a second-degree polynomial was used in the smoothing process. 3/ The left graph was estimated using an Epanechnikov kernel while the graph on the right use a Gaussian kernel. 4/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Appendix 7B: CLSA Two step results. 2010-2013. Other Urban Areas

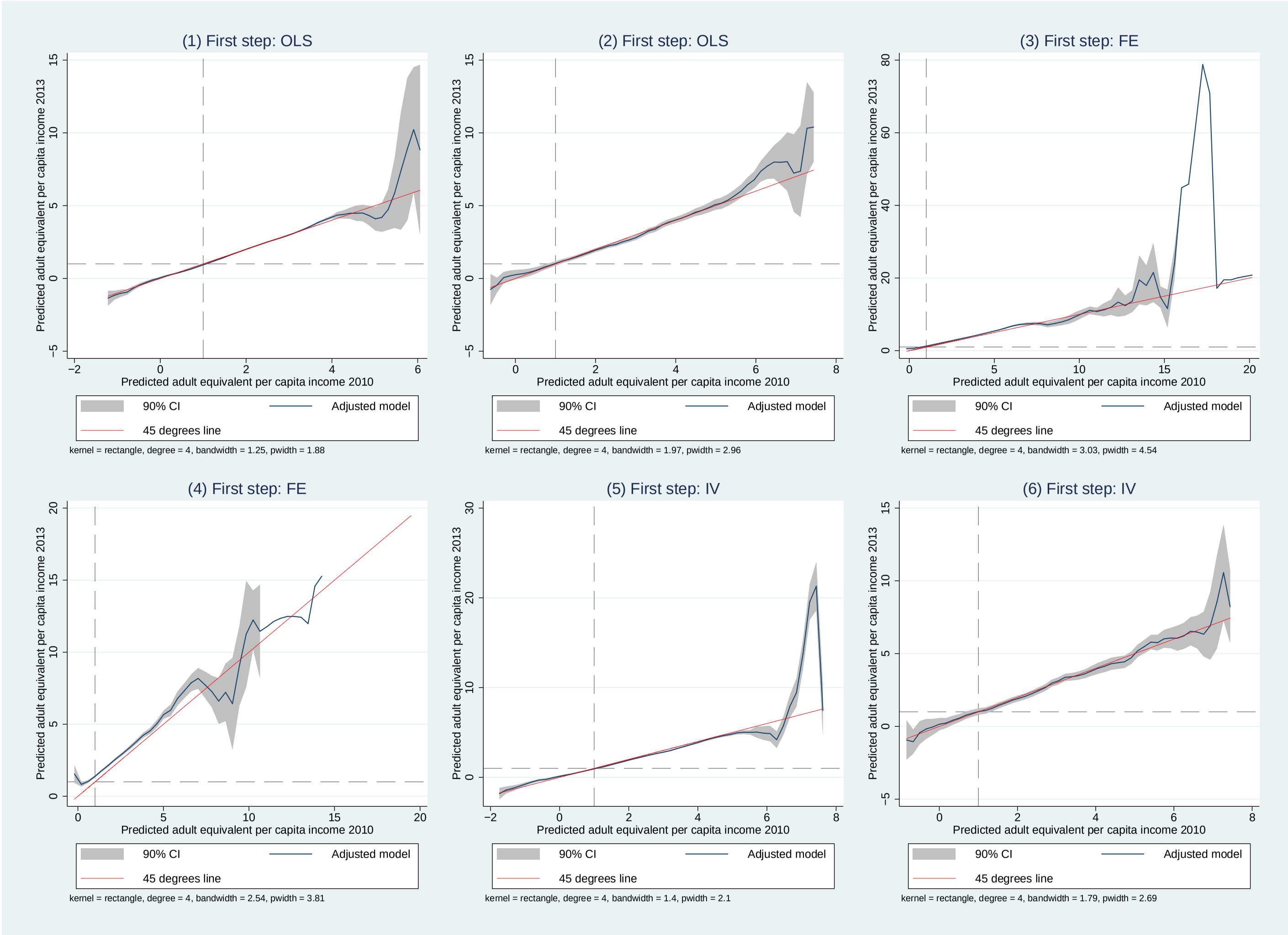
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	OLS (2010)	OLS (2013)	OLS (2010)	OLS (2013)	FE	FE	IV (2010)	IV (2013)	IV (2010)	IV (2013)
Age of head of the household	-0.0247	0.0148	0.1029	0.1327	-0.0886	-0.4065	-0.0254	0.0141	0.2178**	0.1224
	(0.027)	(0.037)	(0.082)	(0.124)	(0.143)	(0.409)	(0.035)	(0.049)	(0.106)	(0.160)
Age of head of the household 2	0.0006**	0.0002	-0.0009	-0.0010	0.0022***	0.0038	0.0008*	0.0003	-0.0023*	-0.0008
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.000)	(0.001)	(0.001)	(0.002)
Education of head of the household	0.1675***	0.1666***	0.1379***	0.1197***	-0.0019	0.0460	0.2458***	0.2147***	0.1571***	0.1956***
	(0.010)	(0.013)	(0.028)	(0.038)	(0.035)	(0.088)	(0.029)	(0.041)	(0.046)	(0.068)
Self-employed head of the household	-0.0408	-0.1995	0.1089	0.0848	-0.1577*	-0.2496	-0.0475	-0.1672	-0.2387	0.2217
	(0.093)	(0.122)	(0.238)	(0.323)	(0.086)	(0.210)	(0.115)	(0.186)	(0.263)	(0.386)
Age of head of household's partner			-0.1657**	-0.1754		0.0636			-0.0620	-0.1396
			(0.084)	(0.133)		(0.256)			(0.103)	(0.166)
Age of head of household's partner 2			0.0022**	0.0019		-0.0004			0.0008	0.0012
			(0.001)	(0.001)		(0.003)			(0.001)	(0.002)
Education of head of household's partner			0.1062***	0.1547***		-0.0245			0.1482***	0.1506*
			(0.030)	(0.045)		(0.054)			(0.056)	(0.083)
Self-employed head of household's partner			-0.1703	-0.2582		-0.1443			-0.3051	-0.2822
			(0.238)	(0.336)		(0.206)			(0.277)	(0.415)
Proportion of members between 0 and 12 years old	-1.2832***	-1.6583***	-1.3440**	-1.9985**	-0.5064*	-0.2661	-1.3542***	-1.7685***	-2.2763***	-2.9639***
	(0.233)	(0.326)	(0.622)	(0.931)	(0.292)	(0.832)	(0.290)	(0.410)	(0.697)	(1.078)
Proportion of members 62+ years old	-0.8596**	-0.2398	-2.7927*	0.4425	0.0400	2.7054	-1.1937*	0.4061	2.2422	2.3360
	(0.343)	(0.406)	(1.538)	(1.927)	(0.403)	(1.823)	(0.657)	(0.596)	(2.853)	(2.941)
Constant	0.9614	-0.2062	-0.0431	1.2409	1.5706	10.1864	-0.6986	-1.0419	-4.4417*	-0.5423
	(1.627)	(2.070)	(2.559)	(3.884)	(5.721)	(14.981)	(0.915)	(1.343)	(2.490)	(4.454)
Observations	1,277	1,277	295	295	2,554	590	869	869	184	184
R-squared	0.32	0.32	0.45	0.42	0.09	0.08	0.34	0.35	0.52	0.43
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
F	10.87	7.858	3.583	2.893	4.143	1.621	.	.	.	.
Number of llave_pers_const					1,350	313				
Year FE					YES	YES				

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ The table presents the first step in the multi-step approach proposed by Adato, Carter and May (2006). Columns 1 and 2 present the results using OLS for estimating the first step for 2010 and 2013 separately. Column 5 show the results, estimating the first step model using all the years available as a pool (2010 and 2013) through an FE data panel model. Columns 7 and 8 estimates the first step model using IV for the years 2010 and 2013 assuming that the education of the head of the household shows reversal causality in respect to the adult equivalent income divided by the poverty line. Education is instrumented with variables of the parents of the head of the household. The other columns includes as additional controls variables of the spouse of the head of the household. 2/ Other urban areas

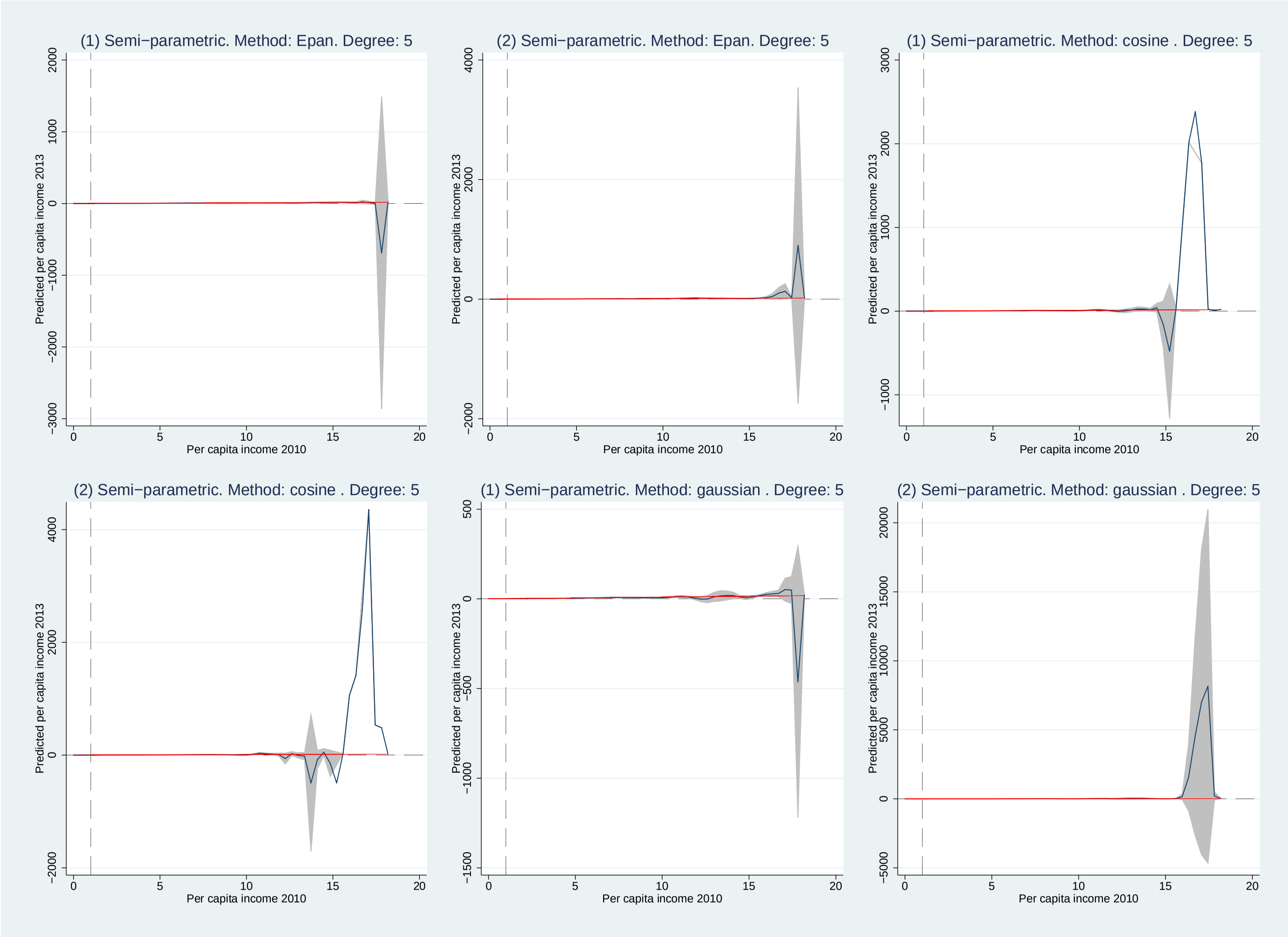
Figure 21: Appendix 7C: Colombian Longitudinal Survey by Universidad los Andes. Second step of the two step method. 2010-2013. Other urban areas



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ The graph presents the second step in the multi-step approach proposed by Adato, Carter and May (2006). The vertical axis variable is the predicted adult equivalent per capita income divided by the poverty line in 2013 and the horizontal axis variable is the prediction for the same variable in 2010. 2/ The non parametric model was estimated using a kernel-weighted local polynomial smoothing regression with a rectangle kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb. 3/ Graphs 1 and 2 use the OLS first step results, while graphs 3 and 4 uses the FE first step results and graphs 5 and 6 uses the first step IV results. Odd graphs use only head of the household controls, while even graphs use head of the household and its spouse controls. 4/ The dashed vertical and horizontal line represents 1 i.e the point where the adult equivalent per capita income is equal to the poverty line. 5/ Other urban areas

Figure 22: Appendix 7D

Vertical axis: Predicted adult equivalent per capita income divided by the poverty line in 2013
CLSA: Semi-parametric results. 2010-2013. Other urban areas



Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ Endogenous variable is the predicted adult equivalent per capita income in 2013, while the controls include, in the parametric part, the same controls included in Table 3, and, in the non-parametric part, the observed adult equivalent per capita income in 2010. 3/ All graphs use a rule of thumb for choosing the bandwidth and a polynomial approximation of the 5th degree. The graphs use cosine, Epanechnikov and gaussian kernel. 4/ The dashed vertical and horizontal lines represent 1 i.e the point where the adult equivalent per capita income is equal to the poverty line.

Appendix 7E: CLSA Semi-parametric results. 2010-2013. Other Urban Areas

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.	Semipar.
Age of head of the household	0.0196	0.0312	0.0134	0.0206	0.0142	0.0267
	(0.019)	(0.038)	(0.018)	(0.031)	(0.018)	(0.032)
Age of head of the household 2	-0.0000	-0.0001	0.0000	-0.0000	0.0000	-0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education of head of the household	0.0599***	0.0498***	0.0568***	0.0491***	0.0568***	0.0505***
	(0.010)	(0.015)	(0.009)	(0.015)	(0.009)	(0.014)
Self-employed head of the household	-0.0930	0.0040	-0.0825	0.0718	-0.0814	0.0537
	(0.090)	(0.151)	(0.090)	(0.148)	(0.090)	(0.148)
Age of head of household's partner		-0.0241		-0.0268		-0.0232
		(0.039)		(0.036)		(0.036)
Age of head of household's partner 2		0.0002		0.0002		0.0002
		(0.000)		(0.000)		(0.000)
Education of head of household's partner		0.0422**		0.0447***		0.0384**
		(0.017)		(0.016)		(0.017)
Self-employed head of household's partner		-0.0928		-0.1332		-0.1391
		(0.149)		(0.136)		(0.138)
Proportion of members between 0 and 12 years old	-0.7896***	-1.1258***	-0.8130***	-1.2194***	-0.7730***	-1.1400***
	(0.206)	(0.425)	(0.202)	(0.403)	(0.203)	(0.403)
Proportion of members 62+ years old	-0.2053	0.3849	-0.2713	-0.0250	-0.2479	0.2254
	(0.331)	(0.800)	(0.298)	(0.628)	(0.302)	(0.705)
Observations	1,606	668	1,606	668	1,606	668
R-squared	0.11	0.18	0.13	0.20	0.12	0.19
City FE	YES	YES	YES	YES	YES	YES
Method	Epan.	Epan.	cosine	cosine	gaussian	gaussian
Degree	5	5	5	5	5	5
F	5226	342.4	3378	94.69	8603	290.3

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Colombian Longitudinal Survey by Universidad los Andes. Note: 1/ Result using the Robinson's semiparametric estimator. 2/ The dependent variable is the adult equivalent per capita income in 2013, while the controls includes in the parametric part the same controls included in Table 3 and in the non-parametric part the observed adult equivalent per capita income in 2010. 3/ It also includes as controls, in other specifications, partner of the head of the household variables. 4/ 14 main cities

Chapter 3. Household income dynamics and income convergence: a new way to measure poverty

Abstract

This paper proposes a new methodology for measuring poverty. The methodology is based on the theoretical framework of asset dynamic poverty traps. The estimation approach follows a two-step methodology. In the first step an asset index -or structural income- is found regressing the household per capita income on the assets that it possesses using a Generalized Linear Model (GLM). In the second step I estimate the dynamics of the structural income. This paper suggests estimating that dynamics using a methodology which classifies the households according to a common growth component of the asset index -or structural income-, along the lines of the literature of "growth convergence clubs" from global economic theory. Such analysis of the structural income is complemented with the analysis of the transitory income -or income shocks- using a panel quantile autoregression method allowing to measure the income shocks persistence. The method is applied to a panel data survey for three main urban areas in Colombia for the years 2007-2010. The results show that 5.8% of the households in these three urban areas are trapped in dynamic structural poverty (i.e, their structural income converges dynamically to an equilibrium below the monetary poverty line) and also that for those households trapped in dynamic structural poverty the transitory income tends to be less persistent compared with the households not trapped in dynamic structural poverty. This implies that households trapped in structural poverty facing a positive transitory income, are temporary lifted out of monetary poverty, but the dynamics of the structural income push them below the monetary poverty line again.

Keywords: Poverty trap, income dynamics, growth convergence, quantile autoregression

JEL classification: I32, O12, O47, B41, C10, C14, C22, C23, C38

1. Introduction

The literature on poverty measurement is rich and has a long tradition. In its current state, it is based on the idea that any poverty measurement should be able to distinguish the structural roots of poverty and investigate the long-term persistence of structural poverty as theorized by Carter and Barrett (2006). This conception of poverty is based on the idea that the more relevant variable when dealing with poverty is not only the level of the current income of households, but also possession of the assets necessary to produce such income. From an applied perspective this implies that the variable to analyse is either an index of productive assets on possession of the household, or the part of the household income that is explained by the asset possession. Carter and Barrett (2006) and Adato, Carter and May (2006) propose to estimate an index of productive assets in possession of the households and analyse the dynamics of said asset index.

Various papers have estimated the existence of poverty traps following the theoretical framework of Carter and Barrett (2006). Their theory helps to determine not only the existence of poverty traps, but also helps identifying the households caught in such a poverty trap, creating in the process a poverty measurement. If a poverty trap exists in the Carter and Barrett (2006) fashion, S-shaped income dynamics with multiple equilibria should be found. Not only that, but the lowest stable equilibrium would be below the monetary poverty line. On the empirical side, the estimation of this kind of poverty traps has been made, mainly, through two-step methods. The first step consists of estimating the structural income -or asset index-. The second step estimates the dynamic relationship between the structural income -or asset index- at two points in time (Carter and Barrett, 2006; Naschold, 2012; Giesbert and Schindler, 2012 and Sandoval, 2019). The most common econometric methods to estimate these poverty traps are non-parametric methods (Lybbert et. al , 2004; Barrett et. al, 2006; Quisumbing and Baulch, 2013; Sandoval, 2019) and semiparametric methods (Kwak and Smith, 2013; Gomez and Lopez, 2013; You, 2014; Sandoval, 2019).

This literature does a remarkable effort on including economic modelling, dynamics and the fundamental causes of poverty in a poverty measure, despite being limited by the econometric methods. Nevertheless, looking for S-shaped dynamics of the structural income (or asset dynamics) has some pitfalls. First, even if a poverty trap exists, an S-shaped curve is only one of the many possible dynamic shapes of the income dynamics. Second, this method assumes that there is only one dynamic path for the average household in the sample, ignoring the distribution of income across households. Third, such models do not take into account the possibility of discrete jumps in the dynamics of the structural income, which may lead to households' structural incomes converging to different dynamic equilibria. Fourth, most of this literature uses only two points on time (two cohorts) for estimating the structural income dynamics, assuming that all households follow the same average dynamic path through time.

Some of the most recent literature has dealt with the first two points exposed here, through the use of some form of conditional or unconditional quantile regression (see Hien, 2011; Kwak and Smith, 2013; Zhou and Turvey, 2015). Indeed, using quantile regression for estimating the dynamics of the structural income improves on previous literature, as it allows for a shape of the dynamics different than an S and models the possibility of multiple dynamic paths for structural income according to the distribution of structural income. Despite such effort, these methodologies have some shortcomings derived from the econometric methodology. First, the quantile regression methods do not allow to interpret the fitted functions for each quantile as the dynamic convergence path for the households belonging to such a quantile. Second, if the quantile to be estimated is chosen low enough, the method will always find a certain proportion of the households trapped in structural poverty. Third, as a consequence of the first two points, these models do not allow to classify which households are indeed trapped in structural poverty.

This paper adds to the current literature on structural poverty traps and poverty measurement in several fronts. First, for estimating the dynamics of structural income, I suggest to follow the method proposed by Phillips and Sul (2007, 2009). This method is based on a neoclassical growth model that allows for heterogeneous technological progress on the income generating process. From here, the authors show that countries (in our case households) can be classified in convergence clubs according to a common growth component of income, using a statistical test developed by them. If the structural income dynamics of some of the growth convergence clubs cross the 45 degree line below the monetary poverty line, those households are indeed trapped in structural dynamic poverty. Second, the method proposed here self-selects the households belonging to each convergence club, instead of classifying them into a certain percentile of the structural income distribution, as the quantile regression methods do. Third, the method itself allows to conclude that the structural incomes of the households belonging to a given club indeed converge to a certain dynamic equilibrium, i.e, the interpretation of the growth convergence clubs is unambiguous. Fourth, this method uses as many observations trough the time dimension as possible, to have a better estimation of the dynamic convergence of the structural income. This contrasts with the methods used in previous literature which in general use only two observations over the time dimension -two survey cohorts-. Fifth, this paper recognizes the possibility of the existence of discrete jumps in the dynamics of structural income and allows the estimation method to capture them accordingly. Sixth, the analysis of structural, income is complemented by the analysis of persistence of the income shocks developed by Arellano, Blundell and Bonhomme (2017). This method uses a panel quantile autoregression to measure the persistence of the income shocks -or transitory income- according to the quantiles of previous income shocks and according to the percentiles of the shocks of the quantile autoregressive process.

The poverty measure proposed here intends to be a complement to the variety of poverty measurements already in existence. As argued by Carter and Barrett (2006) a poverty measurement like the one in this paper, belonging to the fourth generation of poverty measures, distinguishes the structural roots of poverty and investigates the long-term persistence of structural poverty. This characteristic adds to the traditional poverty headcount proposed by Foster, Greer, Thorbecke (1984), to the chronic poverty measurement formalized by Foster (2009) and to the multidimensional poverty index by Alkire and Foster (2011). Indeed, the measure proposed here is a headcount, as it allows to know which households are trapped in structural poverty; it is dynamic, in the same way as the chronic poverty measure, as it takes into account the poverty status in various periods of time; and it is multidimensional, as it is based on the structural income or an asset index. Nevertheless, this work goes beyond these characteristics as it takes into account also the dynamic evolution of the structural income through its growth and the persistence of the structural income across different periods of time.

From a public policy perspective, this measure brings additional information when compared to the traditional poverty headcount (FGT(0)) and the multidimensional poverty measurements which are two of the most popular measurements implemented for evaluating anti poor programs. Mainly, because using the FGT(0) only takes into consideration the poverty status in one period without studying the dynamic evolution of the income. In this respect, households that are never trapped in monetary poverty according to the FGT(0), may have a permanently decreasing structural income, dragging them into monetary poverty in the long run. A similar analysis applies to the multidimensional poverty index. From a theoretical perspective the decomposition of income proposed by Carter and Barrett (2006) into structural income -or asset index- and transitory income -or income shocks- follows the ideas of Friedman's permanent income hypothesis, who recognises that the most important determinant of an individual's consumption is permanent income, (instead of current income) which is in turn a function of the assets possessed by the individual. From an applied perspective the current study uses ground breaking econometric methods which can be applied to a wide variety of econometric problems. Finally, as all the fourth generation of poverty measures, the one in this work is based on economic modelling.

The method proposed here is applied to a panel data survey for three main urban areas in Colombia (the cities of Bogotá, Cali and Bucaramanga) for the years 2007-2010. The growth convergence club method finds the existence of four convergence clubs. These four clubs may be interpreted as a growth convergence club for the households with high structural income, a second and third growth convergence club for the households with middle-high and middle-low structural income and a fourth growth convergence club for the households with low structural income. Furthermore, the dynamics of structural income of the last club converge to a value below the monetary poverty line. I find that 5.8% of the households belong to the lowest convergence club and therefore are trapped in dynamic structural poverty. The analysis of the transitory income for these households trapped in dynamic structural poverty shows that their transitory income tend to be less persistent compared with the transitory income of the households not trapped in dynamic structural poverty. This has some negative and some positive consequences for the households trapped in structural dynamic poverty: on the one hand, for households with positive transitory income this may lift them out of poverty transiently, but the dynamics of the structural income will push them back, below the monetary poverty line in the long run, even if that process takes a long time. On the other hand, for the households with negative transitory income, these negative income shocks tend to disappear quickly and sometimes even reverse through time.

The remainder of this paper is organized as follows: section 2 presents the theoretical framework, section 3 briefly summarizes the literature on poverty trap estimations using household data, section 4 describes the urban Colombian survey used to estimate the method proposed here; section 5 explains in detail the methods used in the paper. Section 6 shows the results of estimating the structural poverty trap presented in section 5 with the data described in section 4. Finally, section 7 briefly presents the conclusions.

2. Theoretical Framework

As Sandoval (2019) points out, the general notion of poverty trap is based on macroeconomics' growth theory, and in particular on the idea of clubs of countries by economic growth. A poverty trap operates as an attractor that does not allow countries to escape from poverty due to its vicious cycle nature (Azariadis and Stachurski, 2005). Arunachalam and Shenoy (2016) classifies macroeconomic poverty traps according to its causes in geographical (Krugman, 1991), imperfect credit (Matsuyama, 2004; Quah, 1996), and coordination failure (Murphy et al., 1989) causes. From a microeconomic perspective the poverty trap theory focuses on households. The more popular approaches include (Sandoval, 2019) models of occupational choice and lack of physical capital (Banerjee and Newman, 1993), lack and quality of human capital accumulation (Galor and Zeira, 1993), malnutrition impeding human capital formation and access to high return activities (Dasgupta and Ray, 1986), and contractual distortions resulting from moral hazard (Mookherjee and Ray, 2002).

The current trend in poverty and inequality measurement recognizes that a good poverty measurement should allow to distinguish the structural roots of poverty[1]

[1] Sandoval (2019) summarizes the Carter and Barrett (2006) idea of an asset based poverty line as a "poverty line that is able to distinguish whether a household is poor due to structural conditions or stochastic shocks. The idea behind this approach is that the productive assets that a household

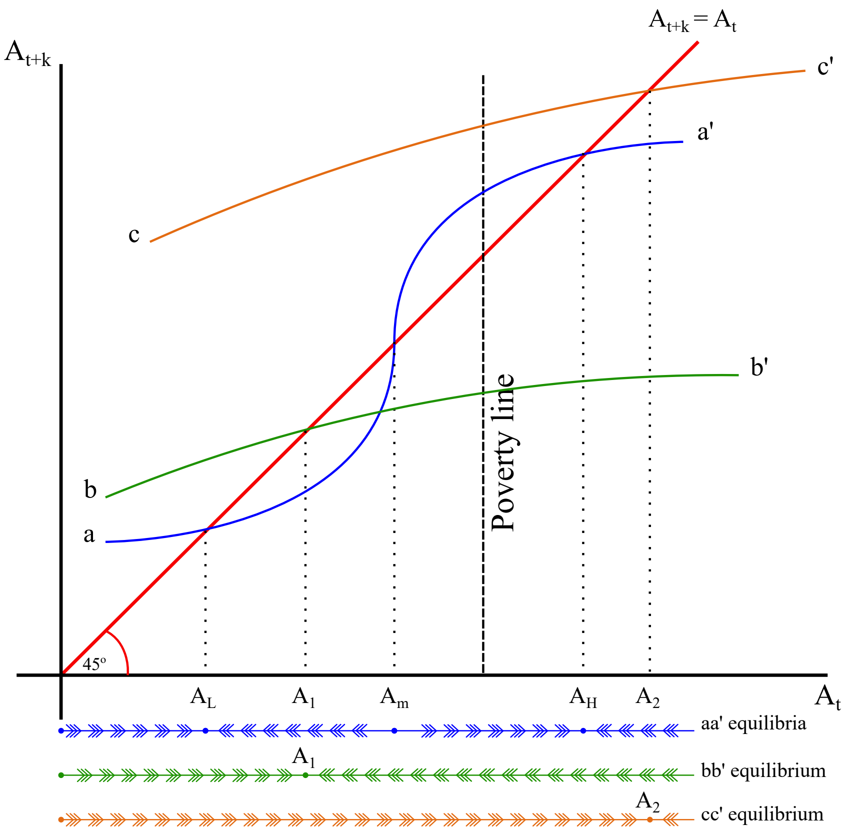
and investigate the long-term persistence of structural poverty (Carter and Barrett, 2006). A dynamic asset poverty line separates those households "caught in a long-term structural poverty trap" from those expected to have a sustained improvement on their structural economic conditions. This section explains the theoretical foundations for the dynamic asset poverty line as developed by Carter and Barrett (2006). Such an asset poverty line is in simple words the "threshold at which accumulation dynamics bifurcate, leading to multiple dynamic welfare equilibria, including the possibility of a poverty trap" (Carter and Barrett, 2006).

Figure 1 shows this situation and the corresponding asset dynamics. It shows clearly how the critical threshold for structural poverty dynamics is not related in any way to the monetary poverty line, which is the most used poverty line. Figure 1 shows on the horizontal axis the household assets in a given initial period t, while the vertical axe shows the household assets in a final period, $t+k$ (Sandoval, 2019). The model proposed by Carter and Barrett (2006) is illustrated in the asset dynamics aa'. Here it can be seen that the dynamic asset poverty line corresponds to the critical threshold A_m, which is the unstable dynamic asset equilibrium and the threshold where the asset accumulation dynamics diverges. A household with wealth above A_m in t would accumulate assets, and ultimately reach a long-term equilibrium asset stock of A_H, generating a steady-state income above the monetary poverty line. In contrast, a household with initial wealth below A_m would give up assets to A_L, and settle at an equilibrium income below the monetary poverty line (Carter and Barrett, 2006).

---

possesses map into the income through certain income generating function, allowing to isolate which part of the income is due to productive assets -structural income- from the portion of the income which is only transitory. Those poverty measurements allow to isolate the structural poverty from the poverty due to stochastic events in a single period".

Figure 1: Different patterns of asset dynamics



Source: Dutta (2015).

Sandoval (2019) argues that the Carter and Barrett (2006) model considers only the possibility of a poverty trap based on the idea of two stable and one unstable equilibria. Nevertheless, this is only one possibility among the infinite possible trajectories of the asset dynamics. In particular, a poverty trap may exist even in the context of a single stable dynamic equilibrium as in Figure 1 (Dutta, 2015). For instance, if the household livelihood function is cc' there is no poverty trap and households converge to the equilibrium point A_2 which is above the monetary poverty line. In contrast, if the household livelihood function is like bb', it would be expected to reach at point A_1, a single steady-state level of equilibrium located below the monetary poverty line (Dutta, 2015). In general, Kwak and Smith (2013) conclude that irrespectively of the number of equilibria there is not reason for the lowest equilibrium in Figure 1 (A_L) to be below a given monetary poverty line.

3. Empirical Literature Review

Many attempts have been made to estimate poverty traps using a wide variety of econometric methods. All the methods described here follow the theoretical perspective of the asset based trap poverty proposed by Carter and Barrett (2006).

The most common method proposed by Barrett et. al (2006) consists of a two-step approach. In the first step an estimation of the so-called "structural income" or "asset based income" is conducted. That is, the part of the income that can be predicted by the productive assets that the household possesses. Such as prediction has been made using OLS models of the household per capita income on physical assets (farm land area, equipment, crops, livestock, inputs for agricultural production, etc.), non physical assets (education, tenure, composition of the work force, etc.) and demographic variables (age composition of the household, gender of the head of the household, etc.) (Barrett et.al, 2006; Adato, Carter and May, 2006; Naschold, 2012; Giesbert and Schindler, 2012; Gomez and Lopez, 2013; You, 2014; and Sandoval, 2019). In the second step, after obtaining such "structural income", an estimation of the dynamics of that variable using a nonparametric method is conducted. Usually a Locally Weighted Scatterplot Smoothing (LOWESS) regression of the "structural income" on its own lags is used for allowing a flexible enough functional form of the estimated dynamics.

In such two-step methods, many other approaches have been used for estimating the first step. Among others, panel data models with fixed effects (Naschold, 2012; Mukasa, 2015; and Sandoval, 2019), panel data models with random effects (Quisumbing and Baulch, 2012) and 3SLS for taking into account the endogeneity between income and educational status (Zhou and Turvey, 2015; and Sandoval, 2019). Other empirical approaches to the estimation of poverty traps include models that do not distinguish among structural income and income shocks, but focus only on estimating the dynamics of the total income using nonparametric methods (Lybbert et.al, 2004; Kwak and Smith, 2011; Maxwell, 2013), parametric methods including GMM (Jalan and Ravallion, 2002; Kwak and Smith, 2013; Mukasa, 2015; Rodriguez and Gonzales, 2004), models considering only labour income using a panel data method which corrects for measurement error (Antman and Mc-Kenzie, 2007) and dynamic data panel methods (Zerfu, 2012; Rodriguez and Gonzales, 2004).

On a different approach Arunachalam and Shenoy (2017) develop their own method for detecting poverty traps based on the idea that if far from a low attractor, the probability of a negative income shock should decrease in a two equilibria environment. Dutta (2015) uses a local polynomial regression for estimating multidimensional poverty traps on the dimensions of illiteracy and undernutrition.

Most of the literature previously mentioned searches for a trap in the way proposed by Carter and Barrett (2006), i.e, those papers look for S-shape dynamics of the "structural income". In case of existence of a poverty trap, two stable equilibria should exist: one below the monetary poverty line and one above the monetary poverty line while there should be an unstable equilibrium between those two. Some other works look for a single equilibrium, assuming that all the households in the database converge to the same equilibrium.

Nevertheless, such methods ignore the possibility of discrete jumps in the dynamics of the structural income, which would lead to the possibility of household structural income converging to different dynamic equilibria, in the same way that was first proposed by the literature on growth convergence clubs in global growth theory.

Only few papers have acknowledge of the possibility of such discrete jumps in the convergence of the income dynamics of the households. Usually, these papers estimate the income dynamics using conditional quantile regression, or some form of unconditional quantile regression. Hien (2011) uses conditional quantile regression to estimate five quantiles of the Vietnamese household income for the years 2004, 2006 and 2008, using as control lagged income. The author finds that households belonging to the lowest income quintile in the last period have a hindered income growth when compared to other less poor households by initial income level.

Kwak and Smith (2013) examine changes in patterns of equilibria over time and across regions, applied to Ethiopian rural regions. The authors revisit incidence of multiple equilibria using nonparametric quantile regression and find that there is a single equilibrium that remains stagnant below the monetary poverty line for households in the lowest 25 percentile of the structural income in the final year.

Zhou and Turvey (2015) quantify the link between agricultural income, caloric in-take, and asset-based poverty in rural China. The models reveal the role that various shocks play in determining the asset dynamics between different income groups using four different asset indexes which cover comprehensive, fixed, productive, and consumable assets through a fourth-degree polynomial function and a three-stage least squares system to account for endogeneity. The authors include quintile dummies according the household asset index in different years. In that way, they estimate the dynamic transition of the households among different asset quintiles. The empirical results do not show evidence of a poverty trap based on multiple equilibria.

The fact that these papers recognize the possible existence of discrete jumps in the dynamics of structural income, jointly with modelling not only the persistence of the poverty status but also the income distribution, force these authors to implement estimation methods accordingly, and can be considered as an improvement with respect to previous literature. Nevertheless, most of them use quantile regression as estimation method which comes with various disadvantages. First, the quantile regression methods do not allow to interpret the fitted functions for each quantile as the dynamic convergence path for the households belonging to such quantile.

When the estimation method for finding dynamic structural poverty traps was first introduced by Adato, Carter and May (2006), they estimated the dynamics of the structural income using the whole sample, implying that the fitted curve can be interpreted as the dynamics of the structural income for the average household in the sample[2]. Nonetheless, this very same interpretation does not translate when using the quantile regression approach for estimating the structural income dynamics. Mainly because for estimating the dynamics of each conditional quantile of interest the whole sample is used, and the econometric interpretation of the fitted dynamics (the fitted line for the Qth quantile) is the function that leaves below such line Q% of the sample conditional on the exogenous variables. In this case that exogenous variable is the lag of the structural income. In other words, the fitted function for the Q quantile cannot be interpreted as the convergence trajectory of the Q quantile of the structural income, but only the fitted line that leaves below Q% of the structural income conditional on the values of the lagged structural income.

[2]Estimating the dynamics of the structural income refers to estimating, for instance, the line aa' in the figure 1 using as dependent variable the structural income at $t + k$ and as regressor the structural income at t.

Second, as the quantile regression method fits a regression of the Q quantile of the endogenous variable conditional on the exogenous variable, it is always possible to find and adjust a model for a quantile low enough such that the adjusted function crosses the 45 degree line below the monetary poverty line. That is to say, if the quantile is chosen low enough, the method will always find a certain proportion of the households trapped in dynamic structural poverty.

Third, as a consequence of the first two points, the models that use quantile regression as estimation method, do not allow to classify which households are really trapped in structural poverty, which should be the final goal of a poverty measurement with this feature.

The present work aims at improving the literature previously mentioned in various aspects. First, the method proposed here allows to clearly identify households that are trapped in structural dynamic poverty, following the method developed by Phillips and Sul (2007, 2009). This method finds income growth convergence clubs among nations -households in this case- and allows the data to self-select the number of growth convergence clubs and the households belonging to each club. If the structural income dynamics of some of the clubs cross the 45 degree line below the monetary poverty line, the households belonging to these clubs are indeed trapped in structural dynamic poverty. Second, the interpretation of the clubs itself is convergence to certain dynamic equilibria, by construction of the method, instead of the fuzzy and unclear interpretation of the quantile regression approach. Third, the analysis of structural income is complemented here by the analysis of the persistence of transitory income -or income shocks- proposed by Arellano, Blundell and Bonhomme (2017). Fourth, this paper does not limit the search of a poverty trap to finding an S shaped structural income dynamics, but recognizes the possibility of the existence of discrete jumps on the dynamics of the structural income.

4. Data

The method proposed here for estimating poverty traps requires at least 4 observations through the time dimension to be implemented. Therefore, this paper uses the panel of the Fedesarrollo Longitudinal Social Survey which follows Colombian urban households. This section discuss the definitions of the variables used in the econometric models, shows descriptive statistics of those variables and discusses technical aspects of the survey. This paper uses the same data and most of the same variables as Sandoval (2019), mainly to keep comparability across the methods in both studies. For this reason, the data section here follows closely the data section in Sandoval (2019). The variable definitions and descriptive statistics are in most cases a reproduction of the text in Sandoval (2019).

4.1 Fedesarrollo Longitudinal Social Survey

The Fedesarrollo Longitudinal Social Survey (FLSS) followed a panel of urban households in Colombia from 2004 to 2010. (Fedesarrollo, 2010). This survey asks yearly info in aspects as dwelling quality, welfare conditions, demographic conditions, health, education and labour market. In 2007 the survey sampled the urban population of Bogotá, Bucaramanga and Cali. Thereafter the universe was expanded to 10 additional main Colombian cities. In 2010 -the last year when the survey was conducted- it was representative for the urban population of 13 cities, approximately for 39.2% of the Colombian population (Sandoval, 2019).

Given the rotating panel design of the FLSS, it is exposed to attrition. According to the technical documentation (Fedesarrollo, 2010), in 2010 only 43.2% (802 households) of the original sample in 2007 is observed in 2010 (Sandoval, 2019). These 802 households constitute the balanced sample for the period 2007-2010. Nevertheless, the effective sample is 684 households as not all the 802 households have information for all the variables to include in the econometric models (Sandoval, 2019). With this in mind, the final sample covers the years 2007-2010 and the cities of Bogotá, Bucaramanga and Cali and it is representative of the population of these three cities (22.3% of the total population in Colombia)[3].

Table 1 shows the basic descriptives of the variables used in the econometric models for the FLSS. First, it is important to note that I am only considering households that did not move across cities between any of the four rounds of the survey, so the

[3]The Phillips and Sul (2007) method requires at least 4 observations across the time dimension for being implemented. Additionally, there are households that are followed from 2004 to 2010. Nevertheless using such sample would greatly reduce the cross section dimension to around 200 households.

distribution of the households in the whole period when the survey was conducted remains the same. 50.3% is located in Bogotá, 45.3% is located in Cali and the remaining 4.4% is living in Bucaramanga. The employment rate among household members[4] decreased yearly in a sustained way from 49.8% in 2007 up to reaching 47.9% in 2010. The age of the head of the household increases by approximately 4 years as is expected between 2007 and 2010, while the education of the head of the household increases only in 0.3 years of education[5] during the same period. This small increment in head of household's education is due, mainly, to the average age of the head of the household. The proportion of members between 0 and 12 years old decreases from 15.8% in 2007 to 13.9% in 2008 up to reaching 11.7% in 2010, which is an expected characteristic in urban populations with low birth rates. Such a characteristic is complemented with an increase in the proportion of members of 62+ years old going from 17.5% in 2007 to 21.8% in 2010 with a sustained decrease rate of one percent point per year. The per capita income adjusted by adult equivalence and standardized by the monetary poverty line[6] increases from 1.9 in 2007 to 2.5 in

[4]To be precise with the working status variables, various definitions should be considered. First of all, the total population can be divided in working age population (people between 12 and 65 years old) and those who do not belong to the working age population (people below 12 years and older than 65 years). The working age population is divided among the occupied individuals (people who declare that they spent most part of the last week working, or people who declare that they spent most of the time during the last week working in his/her own business), the unemployed (people who claim that was looking for a job during the last four weeks and who was able to accept one in case of being offered some but didn't find one) and the inactive (people who declare that is permanent unable to work due to physical condition, people who declare that do not want to find a job or start a business, people who spend most of their time studying, people who haven't tried to find a job in the last four weeks, and those who are not available to take a job in case of being offered). The occupied fraction of the population can be divided in employee, sub-employed, independent worker (including employer) and housekeeping activities. The employment rate is defined as the quotient between the occupied population and the working age population.

[5]Years of education was constructed using two questions. The first asks what is the highest educational attainment of the individual (either complete or incomplete). Those include no education, pre-primary school, primary school, lower secondary education, upper secondary education, technical education, bachelor and graduate school. The second question asks for the highest year reached in said level of educational attainment. Combining these two questions and knowing the typical correlative between educational attainment and years needed to complete said attainments in Colombia, allows to construct the years of education.

[6]The per capita income adjusted by adult equivalence and standardized by the monetary poverty line (Y_{it}) is calculated based on the adult equivalence scale estimated by Muñoz (2014) for Colombia:

$$Y_{it} = \left(\frac{\text{Total Household Income}_{it}}{1+0.7089*(\text{Adults}_{it}-1)+0.6822*\text{Children}_{it}+0.6628*\text{Teenagers}_{it}} \right) \left(\frac{1}{\text{Poverty line}_t} \right)$$

where Total Household Income$_{it}$ is the household income from all sources and all members, for the household i at period t, Adults$_{it}$ is the number of household members 18+ years old for the household i at period t, Children$_{it}$ is the number of household members between 0 and 7 years old

2008 and remains around the same value for the years 2009 and 2010. In other words, the average income of the household changes from 1.9 times the monetary poverty line to 2.5 times the monetary poverty line between the initial and the final year of the survey. Finally, the number of self-employed[7] household members decreases from almost 1.1 in 2007 to 0.9 in 2010 showing a decreasing trend during the whole period.

---

and Teenagers$_{it}$ is the number of household members between 8 and 17 years old. Poverty line$_{it}$ is the urban poverty line for the year t.

[7] A household member is considered self-employed if he declares receiving zero income as wage for being employed in a third party firm and he/she declares working as pawn, on his own business or is an employer. A household member is also considered self-employed if he declares receiving zero income as wage for being employed in a third party firm and he/she declares perceiving income only from its own business, leasing, pensions, other households, government, interest, profit and dividends. Including in this definition as self-employed households who perceive income only from pensions, other households and government is not ideal, but they cannot be separated from the households with other source of income due to how the questions were asked.

Table 1: Descriptive statistics Fedesarrollo Longitudinal Social Survey. 2007-2010. Household level

	(1)	(2)	(3)	(4)	(5)	(6)
	year 2007			year 2008		
VARIABLES	N	mean	sd	N	mean	sd
Bogota	684	0.503	0.500	684	0.503	0.500
Cali	684	0.453	0.498	684	0.453	0.498
Employment rate among household members	684	0.498	0.286	684	0.495	0.295
Male head of the household	684	0.665	0.472	684	0.667	0.472
Age of head of the household	684	52.93	14.03	684	53.91	13.98
Years of Education of head of the household	684	10.89	4.686	684	11.15	4.664
Proportion of members between 0 and 12 years old	684	0.158	0.184	684	0.139	0.179
Proportion of members 62+ years old	684	0.175	0.283	684	0.193	0.303
Percapita income adjusted by adult equivalency and standardized by poverty line	684	1.922	2.331	684	2.508	3.051
Number of self-employed household members	684	1.085	1.063	684	1	0.979

	(1)	(2)	(3)	(4)	(5)	(6)
	year 2009			year 2010		
VARIABLES	N	mean	sd	N	mean	sd
Bogota	684	0.503	0.500	684	0.503	0.500
Cali	684	0.453	0.498	684	0.453	0.498
Employment rate among household members	684	0.488	0.295	684	0.479	0.297
Male head of the household	684	0.642	0.480	684	0.637	0.481
Age of head of the household	684	54.91	13.75	684	55.64	13.91
Years of Education of head of the household	684	11.20	4.686	684	11.19	4.736
Proportion of members between 0 and 12 years old	684	0.133	0.175	684	0.117	0.166
Proportion of members 62+ years old	684	0.205	0.312	684	0.218	0.314
Percapita income adjusted by adult equivalency and standardized by poverty line	684	2.486	2.810	684	2.526	3.215
Number of self-employed household members	684	0.949	0.934	684	0.904	0.963

Source: Author's calculations using Fedesarrollo Longitudinal Social Survey (FLSS). Based on Sandoval (2019)

5. Methodology

This paper follows the two-step approach proposed by Adato, Carter and May (2006) for measuring poverty traps through asset dynamics. In the first step, an asset index -or structural income- is estimated for each household, and in the second step the dynamics of such asset index is estimated.

The observed per capita income of household i in year t, Y_{it}, can be decomposed into structural income, i.e the portion of the income that depends on productive assets (Y_{it}^S), and the portion of the income which is only transitory, i.e income shocks (Y_{it}^T):

$$Y_{it} = Y_{it}^S + Y_{it}^T \quad (Eq.1).$$

Carter and Barrett (2006) and Adato, Carter and May (2006) propose to construct the structural part of the income Y_{it}^S, either, as a unidimensional asset index $\Lambda(A_{it})$ aggregating the productive assets of household i at time t, represented by the vector A_{it}, or finding the part of the total income (Y_{it}) that can be explained by such vector of assets (A_{it}). In this paper I follow the second approach.

When following this method Adato, Carter and May (2006) argue that the identification of an asset poverty line can be done through the estimation of a model that uses as dependent variable some proxy for the livelihood of household i at time t, l_{it}, and as explanatory variables the set of assets in possession of the household during the same period of time (A_{it})

$$l_{it} = f\Big(\sum_j \beta_j A_{ijt}\Big) + \varepsilon_{it} \quad (Eq.2).$$

I measure household livelihood (l_{it}) or material well-being as the household per capita income adjusted by adult equivalence and divided by the monetary poverty line (Y_{it}) of the corresponding year (i.e $l_{it} = Y_{it}$)[8]. Therefore, the dependent variable is equal to one if the household per capita income equals the monetary poverty line.

Most of the previous literature uses some form of linear regression for estimating

[8]See the data section for the precise definition.

the structural part of the income, Y_{it}^S (i.e $f\Big(\sum_j \beta_j A_{ijt}\Big) = \sum_j \beta_j A_{ijt})$. Nevertheless, some studies uses some nonlinear approach to the problem, including quadratic and cubic terms of the j-th asset. If the linear approach is used, the coefficients of the regression relationship, β_j, give the marginal contribution to livelihood of the j different assets. In general, such marginal contribution for a household with J assets is given by

$$\frac{\delta f\Big(\sum_{j=1}^{J} \beta_j A_{ijt}\Big)}{\delta A_{ijt}} \text{ for } j = 1, ..., J$$

Given estimates of the β_j, I can then calculate the fitted value of the regression function, Λ_{it} or equivalently Y_{it}^S, defined as:

$$Y_{it}^S = \Lambda_{it} = E\Big\{f\Big(\sum_j \hat{\beta}_j A_{ijt}\Big)\Big\} \quad (Eq.3).$$

Where $E(.)$ represents the expected value operator.

There are two different interpretations for such estimation. The first interpretation is that Y_{it}^S can be seen as the part of the observed income that can be predicted by the productive assets in possession of the household, i.e the structural income. The second interpretation is that Λ_{it} is an asset index, in which the assets are accompanied by a weight load representing "its marginal contribution to livelihood as given by the estimated coefficients, $\hat{\beta}_j$" (Adato, Carter and May, 2006) and filtered by the function $f(.)$.

Analogously, the income shocks or transitory income can be seen as the residuals of the regression previously described:

$$Y_{it}^T = \hat{\varepsilon}_{it} = Y_{it} - E\Big\{f\Big(\sum_j \hat{\beta}_j A_{ijt}\Big)\Big\} \quad (Eq.4).$$

The remainder of this section explains in detail how the structural income, Y_{it}^S, and the transitory income, Y_{it}^T, will be analysed, and how this allows to create a structural dynamic poverty measure based on assets. The structural income, Y_{it}^S will be studied following the "growth convergence club" classification method proposed by Phillips and Sul (2007, 2009) while the transitory income, Y_{it}^T will be explored following the non-linear dynamic quantile panel method proposed by Arellano, Blundell and Bonhomme (2017).

5.1 Analysis of structural income Y_{it}^S and the existence of poverty traps through household asset dynamics

This subsection explains in detail the growth convergence classification method by Phillips and Sul (2007); The general theoretical method is developed in Phillips and Sul (2007), while its application to growth convergence theory is explained in Phillips and Sul (2009). This paper borrows most of the notation and equations from the later work. The application developed for growth convergence clubs among countries is motivated from a neoclassical growth model with labour augmented technological progress (see the technical appendix of Phillips and Sul, 2009) and time heterogeneous technology by allowing technological progress. Nevertheless, the general statistical method is a classification method for individuals in a panel allowing to categorize them in groups according to the growth of a given variable across time.

Let Y_{it}^S be the observation for individual, country or household i at time t belonging to a panel data set observed across N individuals and T periods of time. Here Y_{it}^S represents the household's structural income. Phillips and Sul (2009) propose to decompose Y_{it}^S as $logY_{it}^S = b_{it} * \mu_t$ where μ_t is a common factor of the structural income among households and b_{it} is an idiosyncratic loading factor that measures the distance of $logY_{it}^S$ to the common component μ_t (Panopoulou and Pantelidis, 2009).

Phillips and Sul (2007) and Phillips and Sul (2009) propose to model the idiosyncratic loading factor b_{it} as the following "relative transition coefficient":

$$h_{it} = \frac{logY_{it}^S}{N^{-1}\sum_{i=1}^{N} logY_{it}^S} = \frac{b_{it}}{N^{-1}\sum_{i=1}^{N} b_{it}} \quad (Eq.5).$$

it is straightforward to show that this quotient eliminates the common component μ_t resulting in a measure of the idiosyncratic component of the structural income of household i at time t, b_{it} in respect with the average of the transitory component across households at time t, $N^{-1}\sum_{i=1}^{N} b_{it}$ (Phillips and Sul, 2009). In Phillips and Sul (2009) words "The variable h_{it} traces out an individual trajectory for each i relative to the average, so we call h_{it} the 'relative transition path'. At the same time, h_{it} measures household i's relative departure from a common steady-state growth path μ_t". As Phillips and Sul (2009) show if the transition behaviour of the structural income of all households is similar asymptotically (when $t \to \infty$), $h_{it} = h_t$ for all i. Moreover, if the structural income of all households tends to the same value when $t \to \infty$, then the numerator and denominator of h_{it} would be the same, and consequently $h_{it} \to 1$, for all i, as $t \to \infty$. This is what Phillips and Sul (2009) call "ultimate growth convergence". Phillips and Sul (2007) show that the conditions $h_{it} \to 1$ and $b_{it} \to b$ imply that under heterogeneity:

$$lim_{t\to\infty} \frac{logY_{it}^S}{logY_{i't}^S} = 1 \text{ for all } i \text{ and } i' \quad (Eq.6).$$

5.1.1 The Log t convergence test

Phillips and Sul (2007) show that if the condition $b_{it} - b_{i't} \to_p 0$ holds, a mean square measure can be developed that eventually may be operationalized as a test. This measure is given by (Phillips and Sul, 2009):

$$H_t = N^{-1} \sum_{i=1}^{N} (h_{it} - 1)^2 \quad (Eq.7).$$

On the one hand, from equation 7, it is clear that if there is convergence each $h_{it} \to 1$ for all i, as $t \to \infty$ and therefore $H_t \to 0$ as $t \to \infty$ (Phillips and Sul, 2007). On the other hand, if there is no convergence Phillips and Sul (2009) argue that there are three possibilities: first, the quantity H_t converges to a constant, second it is bounded but does not converge and third, it diverges. In the case of club convergence Phillips and Sul (2007) show that H_t usually converges to a positive constant.

To formulate a null hypothesis of growth convergence Phillips and Sul (2009) use a semiparametric model for the transition coefficients that allows for heterogeneity over time and across individuals, where under the null of convergence:

$$H_t \sim \frac{C}{L(t)^2 t^{2\alpha}} \quad \text{as} \quad t \to \infty \quad (Eq.8).$$

for some constant $C > 0$. $L(t)$ is a function with the property that $L(t) \to \infty$ as $t \to \infty$ (Phillips and Sul, 2009), α governs the rate at which the cross-section variation over the transition decays to zero over time (Panopoulou and Pantelidis, 2009). Defining $L(t)$ =log t and creating the log ratio log$\frac{H_1}{H_t}$ the following 'log t'

regression model can be estimated:

$$\log\frac{H_1}{H_t} - 2\log(\log t) = a + \gamma\log(t) + u_t, \quad \text{for} \quad t = T_0, ..., T \quad (Eq.9).$$

where H_t follows the definition in 7. Phillips and Sul (2007) show thorough montecarlo simulations that for small samples ($T < 50$) the $r\%$ initial observations across the time dimension should be dropped for the estimators to keep desirable properties in terms of size and power of the test (Panopoulou and Pantelidis, 2009). For this reason the initial observation in 9 is $T_0 = [rT]$.

According to Phillips and Sul (2007) if the null hypothesis of growth convergence is true, then $\hat{\gamma} \xrightarrow{p} 2\alpha$. The key for identifying convergence, club convergence or divergence is the behaviour of the corresponding t-statistic for $\hat{\gamma}$, which should be estimated using HAC standard errors. Let $\{t_{\hat{\gamma}}\}$ be the t-statistic for $\hat{\gamma}$. In the case of $\alpha > 0$ then $\{t_{\hat{\gamma}}\} \rightarrow +\infty$ as $t \rightarrow \infty$. If $\alpha = 0$, $\{t_{\hat{\gamma}}\}$ converges to a standard normal distribution. With this peculiar characteristic in mind, Phillips and Sul (2009) show that "the convergence test then proceeds as a one-sided t-test of $\alpha \geq 0$. Under the alternative of growth divergence or club convergence, the point estimate of γ converges to zero regardless of the true value of α, but its t-statistic diverges to negative infinity, thereby giving the one-sided t-test discriminatory power against alternatives".

Before proceeding with the growth club convergence classification algorithm it is important to discuss how the theoretical background behind this method applies to household income generating functions instead of aggregated production functions for countries. The model developed by Phillips and Sul (2007) is based on the general idea of separating a panel data on a component common to all individuals from an idiosyncratic component particular to each individual. The relative transition concept and its convergence properties as described in equations 5 and 6 are developed as general concepts that apply to any data panel. Furthermore, Phillips and Sul (2007) recognize that this concept may be applied to areas such as economic growth, labour income and stock price factor modelling.

It is important to understand how the neoclassical growth model described in Phillips and Sul (2007) maps into the income generating process of the household. As shown in the technical appendix of Phillips and Sul (2009) the growth model starts with a production function with labour augmented technological progress. In this case such a production function would be the income generating function of the household as a function of capital stock, labour, human capital and the technology (A). In the household case it is reasonable to assume that the household income depends on the corresponding variables at the household level like physical capital stock, years of education of the household (average, total or head of the household), labour volume and other variables like the age composition of the household and work experience. Two variables represent a major concern in respect to the aggregated growth model: the stock of physical capital and the level of technology.

First, the stock of productive physical capital in possession of the household is a variable that usually is not asked in typical household surveys, mainly because most households are presumed to derive their income from being employed in a third party company -specially in urban areas- instead of having their own business. The main exception to this rule would be the households that derive their income from self-employment, which typically would be a low percentage of the urban households. This implies, that even if a household survey asks for the physical capital stock, most of the urban households would report a value of zero. This has two consequences, one practical and one theoretical: first, when estimating the structural income, Y_{it}^S, I will not be able to control for the productive physical capital stock K_{it}. Second, the speed of convergence parameter of the equation describing the evolution of Y_{it}^S depends on the differential between the initial capital stock and the steady-state capital stock of household i. The last point seems to be less important, as the convergence club classification method is still implementable without estimating this convergence parameter. Indeed, Phillips and Sul (2009) use this neoclassical growth model mainly to motivate the use of the econometric technique they propose to the economic growth problem. In this case it is more concerning not to control for the stock of physical productive capital as it may lead to omitted variable bias and inconsistency of $\hat{\beta}_j$ in equation 3.

Second, the level of technology for producing income is assumed to be heterogeneous ($\pi_{it} = \pi_{i0}e^{x_{it}t}$ were π_{it} is the level of technology for household i at time t). This is a fundamental assumption as it allows to represent the structural income as $logY_{it}^S = b_{it} * \mu_t$. When talking about countries this is a reasonable assumption as given different levels of initial technology, slow rates of technological transference and different rates of technological progress would lead to different trajectories of technological progress. Nevertheless, this interpretation does not translate perfectly to the household level, as it is expected that the technology to produce income of different households is relatively similar. The most logical way to assume a different technology on the household income production would be to recognize that employed and self-employed households face different technologies and that there is a difference in technology sophistication according to the particular industry in which the members of the household are currently engaged. Additionally, this difference in technology across industries where distinct households are employed translates into a difference in the technology of the household income generating function.

Finally, as in any neoclassical growth model there is an equation for the evolution of capital stock which is a function of the saving rate, population growth, and the depreciation rate. Here, the mapping of the variables is straightforward compared with the macroeconomic model.

5.1.2 Growth convergence clubs and economic transitions

A detailed analysis of the clustering procedure is given by Phillips and Sul (2007). The steps for implementing the procedure are briefly summarized by Phillips and Sul (2009).

The clustering procedure has four steps. First, order the households according to the amount of final period structural income (Y^S_{iT}). Second, run the log t regression adding one observation at a time, in descending order, starting from the household with the highest income in the final period. Compare all the $N-2$ t statistics. Among the t statistics such that $\{t_k\} > -1.65$ choose the highest one. The corresponding group forms the core convergence group of size k^*. If $\{t_k\} < -1.65$ for the two households with the highest income in the final period, drop the highest one and restart the procedure starting with the second one. Repeat until finding a pair of adjacent observations such that $\{t_k\} > -1.65$. If there is no such pair of observations in the entire sample, conclude that there is no convergence. Third, sieve the data for new club members. Include all households in the sample to the core group, but including one household at a time. Every time a new household is included in the core group, run the log t-test again. If the estimated t-statistic is greater than a given criterion, let's say c^{*}[9], include this new household in the core convergence club, which previous to its inclusion had k^* members according to step 2. Fourth. After forming the core convergence club (if it exists) following the steps 1 to 3, there are two groups; the core convergence club with size k^{**} (after including new members in step 3) and a group of households not converging with the core group with $N - k^{**}$ members for which convergence was rejected in step 3. For this last group of households run the log t-test to see if $\{t_{\hat{\gamma}}\} > -1.65$. If the hypothesis of club convergence is not rejected among this second group, it can be concluded that there are two convergence clubs: the core convergence group formed in first iteration of steps 1 to 3 and the second group formed in step 4. If no convergence is found among all $N - k^{**}$ members of group 2, repeat steps 1 to 3 to see if the second group can be partitioned into convergence clubs. If no k is found when applying step 2 for which $\{t_k\} > -1.65$ among the members of group 2 with size $N - k^{**}$, then those remaining households do not contain a subgroup exhibiting growth convergence behaviour and therefore it can be concluded that they diverge.

After obtaining the convergence clubs, for instance M convergence clubs, Phillips and Sul (2009) propose to check if some convergence clubs may be merged. With this in mind they propose running the log t test merging the two highest convergence clubs, and to keep merging additional clubs (in descending order) as long as the t-statistics are higher than -1,65. Conclude that these convergence clubs form a new single convergence club. Finally, repeat the process but this time starting with the highest club of the remaining convergence clubs, i.e, starting with the club for which the null hypothesis of convergence was rejected in the previous step. Let M^* be the number of final convergence clubs after doing such convergence test among clubs.

[9]The criterion here was developed assuming $t \to \infty$. Phillips and Sul (2007) show using montecarlo simulations that the size of the clustering test -measuring the failure rate of including convergence members in the correct subconvergence club- goes to zero under the null of convergence as $t \to \infty$. On the contrary, in small samples a nominal size of the test of 0.05 usually implies an actual size of 0.2 when r (i.e the proportion of initial periods discarded data) is 0.3. In the same way, the power of the clustering test -the success rate in excluding nonconvergence members from the correct subconvergence club- goes to unity asymptotically regardless of the critical values used. In finite samples, test power is less than unity and, as larger critical values are employed in the selection procedure higher power of the test is found. In the third step, the choice of the sieve criterion c^* is associated with the degree of conservativeness in the clustering method (Phillips and Sul, 2009). Higher c^* implies less risk of including a wrong member of the convergence club. As c^* approaches zero from below, the sieve condition becomes more conservative. When t is small, the sieve criterion c^* can be set to zero to ensure that it is highly conservative. Therefore, in this case as t=4, I am setting c^*=0.

5.2 Analysis of the transitory income Y_{it}^T and the persistence of income shocks

Arellano, Blundell and Bonhomme (2017) propose a method to analyse transitory income and the persistence of income shocks based on a quantile-based panel data framework. The first step is to construct Y_{it}^T as the residuals from regressing some kind of household income on a set of demographics. Such method corresponds to the definition given in equation 4. The equations and notation in this subsection follow closely Arellano, Blundell and Bonhomme (2017).

5.2.1 Quantile-based panel data model

Arellano, Blundell and Bonhomme (2017) propose to decompose Y_{it}^T in two additively separable components. The first one represents the part of the shocks that persist through time (η_{it}) and the second one represents a random component (υ_{it}) with zero mean, independent across the time dimension and independent of η_{is} for all s[10]:

$$Y_{it}^T = \eta_{it} + \upsilon_{it}, \quad i = 1, ..., N, \ t = 1, ..., T \quad (Eq.10).$$

Arellano, Blundell and Bonhomme (2017) assume that η_{it} follows a general first-order Markov process. The general idea of this method is modelling the persistence of the quantile of the income shocks (Y_{it}^T). In other words, find out if there is a different pattern in the persistence of the income shocks, Y_{it}^T according to the quantile of the distribution they belong to. This is a relevant question as it would allow to distinguish, for instance if negative income shocks (income shocks in the lowest quantiles of the distribution) are more or less persistent in comparison with positive income shocks (income shocks in the highest quantiles of the distribution). For this reason, Arellano, Blundell and Bonhomme (2017) proceed defining the τth conditional quantile of the persistent component in t, η_{it} given (or conditional on) the lag of the persistent component $\eta_{i,t-1}$ as $Q_t(\eta_{i,t-1}, \tau)$ which, as any quantile function, is defined for every $\tau \in (0, 1)$. Combining these ideas, the conditional quantile function can be written as an autoregressive process as:

[10] Arellano, Blundell and Bonhomme (2017) claim that as in any regression model υ_{it} represents a mixture of innovations and measurement error.

$$\eta_{it} = Q_t(\eta_{i,t-1}, \omega_{it}), \quad (\omega_{it}|\eta_{i,t-1}, \eta_{i,t-2}, ...) \sim \text{Uniform}(0,1), \quad t = 2, ..., T \quad (Eq.11).$$

Intuitively this function describes the conditional quantile function (η_{it}) in t as a function of the conditional quantile function in $t-1$ ($\eta_{i,t-1}$) and a shock ω_{it} to the quantile autoregressive process. Arellano, Blundell and Bonhomme (2017) show that for identifying the corresponding parameters to estimate in the process it is only required that η_{it} shows some form of dependence over time.

5.2.2 Nonlinear dynamics

The equations 10 and 11 describe the autoregressive process assumed by Arellano, Blundell and Bonhomme (2017). Furthermore, as the purpose of the exercise is to determine the persistence of income shocks according top the quantile they belong to, it is important to allow for a flexible functional form in the persistence of the autorregresive process. In particular, the model described so far allows for nonlinear dynamics of the transitory income. Arellano, Blundell and Bonhomme (2017) assert that this model "focuses on the ability of this specification to capture nonlinear persistence, and general forms of conditional heteroscedasticity". Following this idea, Arellano, Blundell and Bonhomme (2017) propose to measure the non-linear autoregressive persistence as:

$$\rho_t(\eta_{i,t-1}, \tau) = \frac{\delta Q_t(\eta_{i,t-1}, \tau)}{\delta \eta_{i,t-1}}, \quad \rho_t(\tau) = E\left[\frac{\delta Q_t(\eta_{i,t-1}, \tau)}{\delta \eta_{i,t-1}}\right] \quad (Eq.12).$$

where $\delta Q_t / \delta \eta_{i,t-1}$ is the derivative of Q_t with respect to $\eta_{i,t-1}$ and taking the expected value with respect to the distribution of $\eta_{i,t-1}$ in the right hand side equation.

Given the definitions in equation 11, $\rho_t(\eta_{i,t-1}, \tau)$ may be interpreted as the persistence of $\eta_{i,t-1}$ when it is hit by a shock in period t (ω_{it}), where ω_{it} corresponds to the τth quantile of the distribution of the shocks to the conditional quantile autoregressive process (Arellano, Blundell and Bonhomme, 2017). The right hand side equation in 12 ($\rho_t(\tau)$) can be interpreted as the average persistence across the distribution of

η. Accordingly, the persistence of the histories of $\eta_{i,t-1}$ depends not only on its own lags, but also on whether the shock ω_{it} is positive or negative, and on its scale.

So far, Arellano, Blundell and Bonhomme (2017) are still very general about the form of the model they propose. The next equation proposed by them gives a particular form for the conditional quantile autoregressive process:

$$Q_t(\eta_{i,t-1},\tau) = \psi_t(\tau) + \phi_t(\tau)' h(\eta_{i,t-1}) \quad (Eq.13).$$

where h is a polynomial function. Taking the derivatives in equation 13 with respect to $\eta_{i,t-1}$ and its expected value one can find the persistence and average persistence associated with this polynomial function:

$$\rho_t(\eta_{i,t-1},\tau) = \phi_t(\tau)' \frac{\delta h(\eta_{i,t-1})}{\delta \eta_{i,t-1}}, \quad \rho_t(\tau) = \phi_t(\tau)' E\left[\frac{\delta h(\eta_{i,t-1})}{\delta \eta_{i,t-1}}\right]$$

Traditional quantile autoregressive models consider only the case where the autoregressive coefficients are independent of the quantiles where they are estimated (Koenker and Xiao, 2006). In this context, this would imply ϕ_t instead of $\phi_t(\tau)$. Koenker and Xiao (2006) develop a modification of this allowing the autoregressive coefficients to vary with the quantiles ($\phi_t(\tau)$ with $\tau \in [0,1]$). Arellano, Blundell and Bonhomme (2017) argue that this allows "shocks to affect the persistence of $\eta_{i,t-1}$ in a flexible way".

6. Results

6.1 Results of the analysis of the structural income Y_{it}^S and the existence of poverty traps

This section shows the results of the methods for estimating poverty traps and analysing the household income shocks proposed in section 5. Here, I am using as l_{it} (the measure of livelihood or material well-being) the household per capita income adjusted by adult equivalence and divided by the monetary poverty line of the corresponding year.

Additionally, a Generalized Linear Model (GLM) with a Gamma distribution function and a logarithmic link function was used for obtaining the structural income Y_{it}^S. This is due to some useful properties of said model, mainly:

$$Y_{it}^S = E(Y_{it}) = E(l_{it}) = e^{\beta_0 + \sum_j \beta_j A_{ijt}}$$

Where e represents the exponential function and β_0 is the equivalent to the intercept in a linear model.

First, assuming a gamma-distributed dependant variable is the right choice when the dependant variable is real valued in the range $[0, \infty)$, which is the case when modelling the household per capita income as, by definition, it cannot take negative values. Second, by the properties of the gamma distribution, the expected value of the random variable is proportional to its variance. In this case it implies that households with a low expected household per capita income, should also expect a low variability in the observed values of the per capita income. Conversely, households with a large expected value of the per capita income, should also expect a large variability among the observed values. This is the typical behaviour on the observed distribution of the household per capita income, where the distribution usually shows positive skewness with a longer right tail, reflecting higher variability between higher values of the household per capita income. Third, the application of the club convergence algorithm by Phillips and Sul (2007) requires positive values of the structural income Y_{it}^S, which are guaranteed by the gamma model[11]. Finally, different from its closest counterpart, the log-linear model, the gamma model allows to recover the level of the household per capita income directly without intermediate corrections as those required by the log-linear model.

Table 2 and appendix 1 show the results of the regression proposed in equation 2 using a Generalized Linear Model assuming a Gamma distribution of the house-

[11] Other model specifications were fitted for estimating the structural income Y_{it}^S. Initially a linear model was estimated using OLS. Nevertheless, this option was discarded as it produces a considerable portion of negative values among the structural income (i.e $Y_{it}^S < 0$). This result not only hampers the implementation of the Phillips and Sul (2007) method, but it is also counter intuitive: the structural income is the part of income that depends on structural characteristics of the household, which, in the worst case scenario may lead to a zero per capita income. A log-linear model for the household per capita income was also considered. In this case, it was ruled out due to its difficulty in recovering the levels of the dependant variable, a mandatory requirement when implementing the Phillips and Sul (2007) methodology.

hold per capita income adjusted by adult equivalence and divided by the monetary poverty line with a logarithmic link function. The regressors are a set of demographic variables of the head of the household, a set of human capital variables like years of education and age as a proxy for work experience, variables describing the labour market situation of the household and fixed effects by city. Table 2 shows the marginal effects of the model (given the non-linearity of its nature) and appendix 1 presents the estimated coefficients. As expected the years of education of the head of the household and the employment rate among household members have a positive and significant effect on the expected household per capita income, while the proportion of household members between 0 and 12 years old has a negative and significant effect, as such members do not participate in the labour market. These three variables have a statistically significant and stable effect (in terms of direction) during all the periods.

Of the remaining variables, households with a male head of the household have a positive and significant effect in 2007 and 2009 on household income, while on the other two years seem to have no significant effect, despite of conserving the same sign of the coefficients. The proportion of members 62+ years old have a negative and significant effect on the expected household per capita income in 2007 only. On the other years, it does not have any statistically significant effect on the expected household per capita income while the sign of the effect becomes positive. Furthermore, the expected sign of the effect of this explanatory variable on the household per capita income is not clear: if the 62+ years old member is a pensioner with a pension above the average household income the sign would be positive. On the contrary, if the old household member did not reach a pension, or, he reached it but is below the average household income, its effect would be negative. The same analysis applies for the number of self-employed household members, as their effect on the per capita household income depends on what is the income of the self-employed in comparison with the average income of the other members of the household.

The age and age square of the head of the household were included to capture the positive but decreasing effect of the head of the household tenure on the household income. In three out of the fours years the effect of the linear term is positive, while the effect of the squared term is negative as expected. Fixed effects by city were included to capture regional effects. Nevertheless, none of these variables have a statistically significant effect on the expected value of the per capita household income

Figure 2 shows the kernel density estimates of the structural household income (Y_{it}^S), or in other words, the predicted household per capita income adjusted by adult equivalence an divided by the monetary poverty line. The density shows a little displacement year by year to the right, indicating that there has been an increase in the structural income of the households trough time. Figure 3 shows the scatter plot between the structural income of the year 2007 and 2010 ($Y_{i,2007}^S$ vs $Y_{i,2010}^S$). Again, as expected, there is a positive correlation between the variable in these two years as this is the portion of the income that depends on structural characteristics of the household and that are not expected to fluctuate greatly year by year.

Table 2: Fedesarrollo Longitudinal Social Survey Estimation of the Household Per capita Income Adjusted by Adult Equivalence and Divided by the Monetary Poverty Line (l_{it})

Marginal Effects of a Generalized Linear Model Distribution: Gamma. Link Function: Logarithmic. 2007-2010. Main Urban Areas

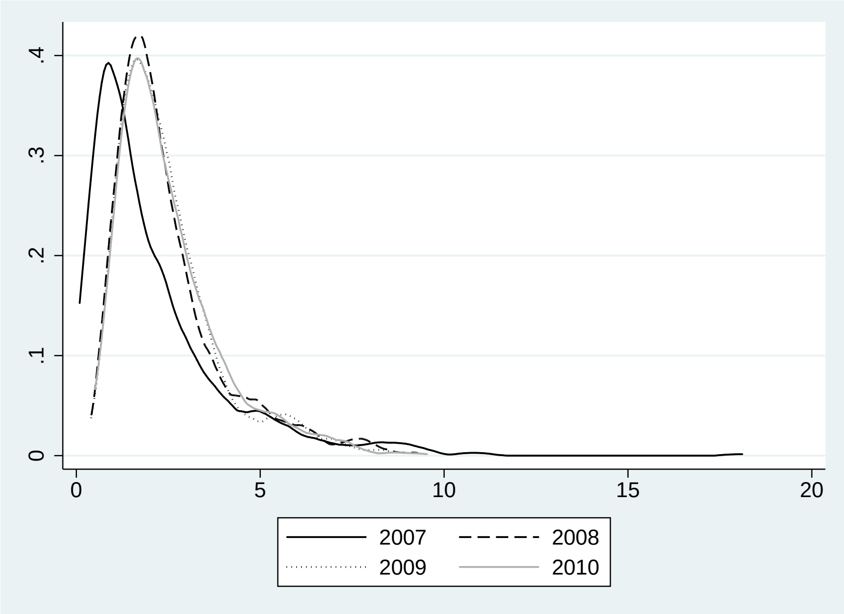
	(1)	(2)	(3)	(4)
VARIABLES	2007	2008	2009	2010
Male head of the household	0.3661***	0.0515	0.3000**	0.2094
	(0.109)	(0.158)	(0.144)	(0.173)
Years of Education of head of the household	0.1181***	0.2034***	0.1652***	0.1803***
	(0.012)	(0.017)	(0.016)	(0.019)
Age of head of the household	0.0262	0.0436	-0.0026	0.0338
	(0.026)	(0.037)	(0.036)	(0.042)
Age of head of the household 2	-0.0000	-0.0001	0.0003	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Employment rate among household members	3.4503***	1.9039***	1.7930***	1.9719***
	(0.281)	(0.292)	(0.279)	(0.326)
Proportion of members between 0 and 12 years old	-1.1667***	-1.7227***	-2.7919***	-2.6004***
	(0.338)	(0.474)	(0.461)	(0.575)
Proportion of members 62+ years old	-1.1433***	0.1665	0.1027	0.4389
	(0.270)	(0.359)	(0.336)	(0.381)
Number of self-employed household members	0.2094***	-0.0157	0.0158	0.1213
	(0.064)	(0.088)	(0.084)	(0.097)
Bogota	0.0617	0.2114	0.1828	0.4137
	(0.268)	(0.367)	(0.348)	(0.417)
Cali	-0.0012	0.4363	0.2601	0.3387
	(0.269)	(0.380)	(0.356)	(0.425)
Observations	684	684	684	684
F	.	.	.	.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

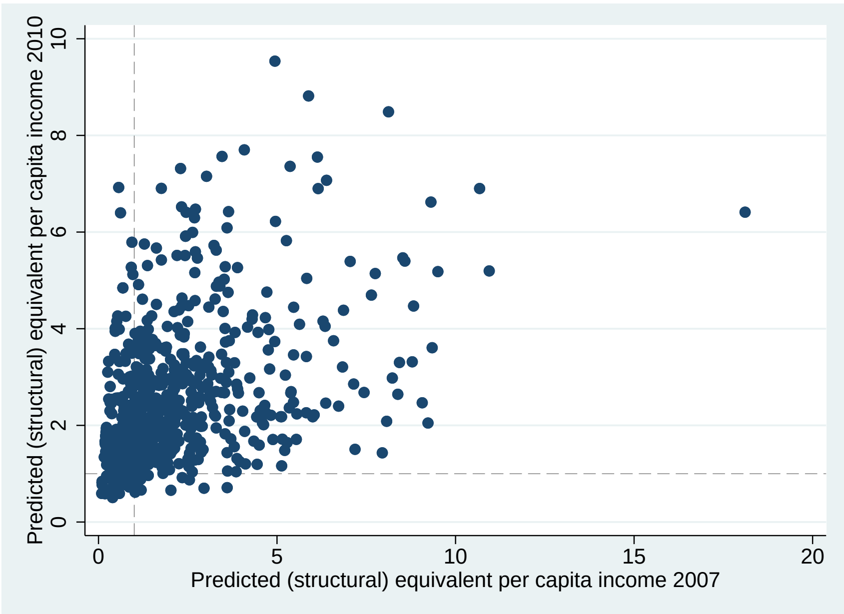
Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ The table presents the first step in the two-step approach proposed by Adato, Carter and May (2006). 2/ Columns 1-4 present the marginal effects of the control variables using a Generalized Linear Model with a Gamma distribution and a Logarithmic link function. 3/ Estimations for all the years available (2007, 2008, 2009, 2010). 4/ For the precise definition of the variables used here see section 4 (data section). 5/ A dummy variable for the city of Bucaramanga is left out

Figure 2: Kernel density estimates of the Structural (or predicted) household income (Y_{it}^{S}) 2007-2010



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS) for the years 2007-2010. Note: 1./ Kernel density estimates of the structural household income (Y_{it}^{S}) using epanechnikov kernel and data driven bandwidth selection method. 2/ The structural household income (Y_{it}^{S}) was estimated using the coefficients of the Generalized Linear Model in Table 2.

Figure 3: Scatter plot of the Structural (or predicted) household income (Y_{it}^{S}) 2007 vs 2010



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS) for the years 2007 (horizontal axe) and 2010 (vertical axe). Note: 1./ Scatter plot of the structural household income (Y_{it}^{S}). 2./ The structural household income (Y_{it}^{S}) was estimated using the coefficients of the Generalized Linear Model in Table 2.

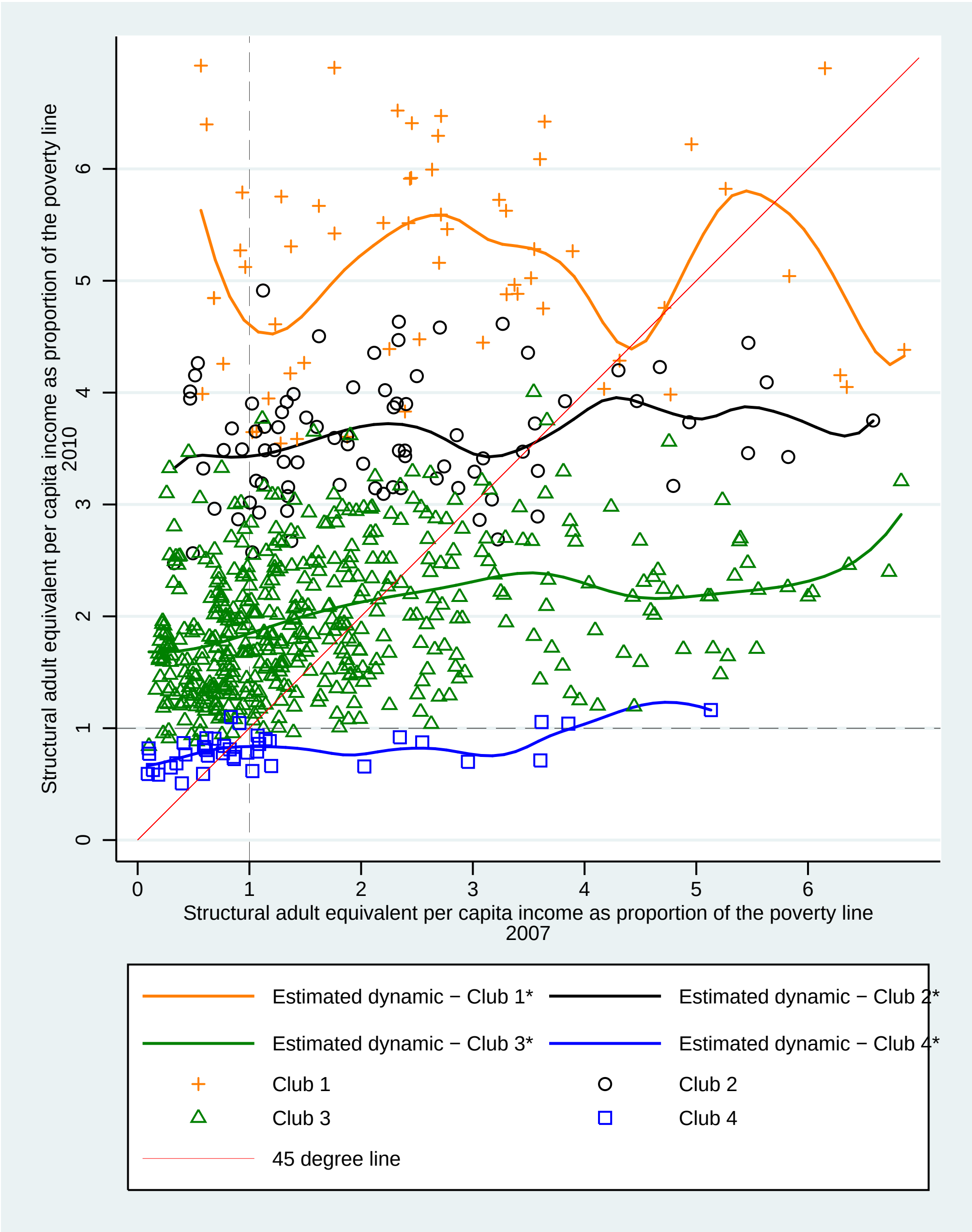
Figure 4 and appendix 2 show the results of applying the four-step algorithm detailed in section 5.1.2 to the structural (or predicted) household income (Y_{it}^S). Various details of such outcome deserve further comment.

First, the result of applying the Phillips and Sul (2007) algorithm is a classification of all the households in the database on different groups or convergence clubs. Therefore, each household belongs to only one growth convergence club across the time dimension. Second, in this particular case the algorithm found six growth convergence clubs for the household structural income in the database. Third, appendix 2 shows the scatter plot between the structural (or predicted) income (Y_{it}^S) on the last year I have data for ($Y_{i,2010}^S$) on the vertical axis, versus the same variable on the first year for which I have data ($Y_{i,2007}^S$) but in this case distinguishing to what growth convergence club each household belongs to. A non-parametric regression was run in the way proposed by Adato, Carter and May (2006) between the variables on the two axes for every club separately. From here, I can conclude that households classified on the convergence club six are trapped in dynamic structural poverty, as the estimated dynamics of the structural income crosses the 45 degrees line below the poverty line (the dotted line, which represents one). Fourth, after applying the log t test among convergence clubs as suggested by Phillips and Sul (2009), the algorithm find a final number of four convergence clubs: *club one**, including households converging to a high structural income, *club two**, composed by households converging to a medium-high structural income, *club three**, comprised by households converging to a medium-low structural income and *club* $four^*$ including the households belonging to the club six of the original algorithm and being trapped in structural poverty.

Figure 4 shows the estimated dynamics of the structural income for these four merged clubs. *Club one** seems to have a dynamic convergence to a structural per capita household income around 4.5 times the monetary poverty line; *club two** converges dynamically to a structural per capita household income of around 3.5 times the monetary poverty line; *club three** converges dynamically to a structural per capita income of around 2.2 times the monetary poverty line and *club* $four^*$ converges to a structural income of around 0.8 times the poverty line. These conclusions can be drawn observing where the adjusted non-parametric line on the scatter plot of each club crosses the 45 degree line. In conclusion, all the households belonging to *club* $four^*$ are trapped in dynamic structural poverty.

Furthermore, the outcome presented here is in principle more robust than any of the results in previous papers using only the method proposed by Adato, Carter and May (2006). This is due to the fact that the method proposed here uses all the data available when finding the convergence clubs. In contrast, the nonparametric regression proposed by Adato, Carter and May (2006) only uses the first and last observations of the data, even when they have more than two observations, as its ultimate porpouse is only to estimate the point where the income dynamics function of the average household crosses the 45 degrees line.

Figure 4: Club convergence result on the Structural (or predicted) income (Y_{it}^S) 2007 vs 2010



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS) for the years 2007 (horizontal axe) and 2010 (vertical axe). Note: 1/ The graph presents the second step in the multi-step approach proposed by Adato, Carter and May (2006) using the convergence club classification method by Phillips and Sul (2007). The vertical axis variable is the structural income (Y_{it}^S) in 2010 and the horizontal axis variable is the structural income (Y_{it}^S) in 2007. 2/ The adjusted model is a nonparametric model estimated using a kernel-weighted local polynomial smoothing regression with a rectangle kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb for each club separately. 3/ The dashed vertical and horizontal line represent 1, i.e, the point where the structural income (Y_{it}^S) is equal to the monetary poverty line. 4/ The Phillips and Sul (2007) algorithm found 6 convergence clubs. Those clubs were merged into for clubs following the results of the log t test. *Club one** has a dynamic convergence to a structural per capita household income around 4.5 times the monetary poverty line; *club two** converges dynamically to a structural per capita household income of around 3.5 times the monetary poverty line; *club three** converges dynamically to a structural per capita income of around 2.2 times the monetary poverty line and *club four** converges to a structural income of around 0.8 times the poverty line. 5/ The households belonging to *club four** are trapped in dynamic structural poverty.

Figure 4 also elucidates how the club convergence algorithm by Phillips and Sul (2007) operates. For instance, households belonging to the growth convergence $club\ four^*$ are households that have a very low structural income (Y_{it}^S) in the final year (2010), typically below 1.2 times the monetary poverty line. Nevertheless these households come from a variety of income backgrounds in the initial year (2007). Some of them had a relatively high income in 2007 (between 2 and 5 times the poverty line), but suffered from negative growth rates of structural income, and therefore end up in structural poverty in the final year (2010) and classified as dynamic structurally poor along the whole period. Some other households are in structural poverty in both years, had a very low or null growth of the structural income and as a consequence are also trapped in dynamic structural poverty. It is remarkable that most of the households in structural poverty in both years are classified as dynamically structurally poor. Nevertheless, there is also a portion of the households that are in structural poverty in both years but are not classified as structurally poor. There is a final set of households that are out of structural poverty in both years but are classified as being trapped in dynamic structural poverty. This result is probably a consequence of a very negative growth rate of the structural income on the years not shown in the graph (2008 and 2009).

The convergence *clubs one**, *two** and *three** have two important characteristics: First, most of the households classified on these convergence clubs are not in structural poverty in both years shown here (2007 and 2010), with some exceptions on households classified among *club three**. Second, the majority of the households belonging to these clubs which are in structural poverty in 2007 increased their structural income and leave structural poverty in 2010, implying that they have faced positive growth of the structural income between these two periods and showing that the algorithm is doing properly the work it was intended for.

Table 3 shows the evolution of different poverty rates through time for the balanced sample used in the econometric models. Row 1 shows the poverty headcount (FGT(0) by Foster, Greer, Thorbecke, 1984) which estimates the percentage of households with a per capita income adjusted by adult equivalence below the monetary poverty line. Rows 2-4 estimate the number of households who are chronically poor. This method counts the number of periods that a household is poor and then according to a predetermined threshold, a household is classified as chronically poor or not.

The results are presented for the cases when a household is considered chronically poor if it is in monetary poverty in two periods or more, three periods or more or four periods out of four. This kind of measure was proposed and used by Foster (2009), Levy (1977) and Duncan and Rodgers (1991).

Row one of table 3 shows the poverty headcount (FGT(0)) for the observed income (Y_{it}). The proportion of households in monetary poverty plummeted from 2007 to 2008 from 35% to 25%, falling another percentage point the next year up to around 24% where it stagnated around the same value in 2010. Rows two to four show the result of estimating the chronic poverty measurement explained previously using the structural income. Row two estimates the number of households considered to be chronically poor, if they are classified as chronically poor when have been in poverty at least two periods (out of four). In this case 9% of households have been in such situation. When the criteria is strengthened to consider a household chronically poor if it has been in poverty for at least three periods, the chronic poverty rate decreases to 4.5% (Row three). If considering chronically poor the households who have been in poverty four periods the chronic poverty rate is reduced to 3% (Row four).

Finally, and most importantly, the row five shows the poverty rate estimated using the method proposed here. From a theoretical perspective, those are the households trapped in dynamic structural poverty, while, from an econometric perspective, those are the households belonging to the merged convergence $club\ four^{*}$ according to the method proposed by Phillips and Sul (2007). In this case, 5.8% of the households (out of 684) are trapped in dynamic structural poverty.

Besides the poverty rate trends estimated in table 3, it is important to compare how these poverty indices relate to each other. In particular, it is important to understand how the poverty measurement proposed here contrasts with traditional poverty rates.

Table 3: Poverty rate estimates

Poverty Rate	2007	2008	2009	2010	Obs.
1. Poverty rate for the observed income (FGT(0))	35.67	25.58	23.83	24.42	684
2. Chronically poor if poor for two or more periods (Structural income)	8.918	8.918	8.918	8.918	684
3. Chronically poor if poor for three or more periods (Structural income)	4.532	4.532	4.532	4.532	684
4. Chronically poor if poor for four periods (Structural income)	2.924	2.924	2.924	2.924	684
5. Dynamic structurally poor (trapped in structural poverty)	5.848	5.848	5.848	5.848	684

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Row 1 shows the monetary poverty headcount proposed by Foster, Greer, Thorbecke (1984) for the observed income (Y_{it}). 2/ Rows 2-4 shows the chronic poverty measurement proposed by Foster (2009) assuming that households are considered chronically poor, if they fall in poverty in different number of periods using the structural income (Y_{it}^S). 3/ Row 5 shows the poverty rate proposed here, i.e the proportion of households trapped in dynamic structural poverty.

Table 4 compares the poverty status of the households in each year according to different traditional poverty measurements (in the rows) and the dynamic structural poverty status (in the columns). Row one estimates the traditional poverty headcount (FGT(0)) on the observed income (Y_{it}), year by year and compares it with the poverty measure proposed here, i.e, the poverty rate according to the dynamic structural measurement. In 2007, 3.6% of the households are poor according to both measures, while 32% of the households are poor according to the traditional monetary measurement (FGT(0)) only. This can be interpreted as 32% of the households being non structural but transiently poor, having a per capita income below the monetary poverty line in 2007. Likewise, 2.1% of households are dynamically structurally poor, but not poor according to the traditional monetary poverty rate (FGT(0)). In other words, 2.1% of the households have a per capita income above the monetary poverty line, but are considered to be trapped in dynamic structural poverty. These households are transiently out of monetary poverty, but their condition of being structurally and dynamically poor eventually will drag them to monetary poverty. The distribution of those poverty statuses remain relatively constant across the whole period.

Row number two cross tabulates the measure of chronic poverty explained before with the dynamic structural poverty measurement proposed here. In the case when a household is considered chronically poor if its structural income (Y_{it}^S) is below the monetary poverty line for 2 periods or more, 89% of the households are not poor according to any of the two poverty measurements, while 4.2% of the households are poor according to both measures. Around 1.6% are trapped in dynamic structural poverty but are not considered chronically poor and 4.6% of the households are chronically poor but not trapped in dynamic structural poverty. When the concept of chronically poor is more restrictive to include only households with structural income below the monetary poverty line during three or more periods (row three), 93% of the households are not poor according to any of the two poverty measures while 1.3% of the households are chronically poor according to the traditional measure but not dynamic structurally trapped in poverty. 2.6% of the households are trapped in structural dynamic poverty but not chronically poor, while 3.2% of households are trapped in structural dynamic poverty and are chronically poor. When a household is considered chronically poor if its structural income is below the monetary poverty line in the four periods, 93% of the households are not poor according to any of the two poverty measures, while 0.7% are poor according to the chronically measure but not poor according to the dynamic structural measure. 3.6% of the households are not poor according to the chronic poverty measure but are trapped in structural dynamic poverty. Finally, 2.1% of the households are chronically poor and trapped in dynamic structural poverty.

This analysis is important as it shows the difference between the chronic poverty measurement and the dynamic structural poverty measure: those numbers differ because the dynamic structural measurement also takes into account the persistence of the structural poverty status and does not count the number of periods in structural poverty as independent events.

Table 4: Poverty rate estimates comparison

Columns: Households trapped in dynamic structural poverty. Rows: Traditional poverty measurements

		Poverty status according to dynamic structural poverty measurement							
Poverty Rate	Status	2007		2008		2009		2010	
		Non poor	Poor	Non poor	Poor	Non poor	Poor	Non poor	Poor
1. Poverty rate for the observed income (FGT(0))	Not Poor	62.13	2.193	72.08	2.339	73.68	2.485	73.39	2.193
	Poor	32.02	3.655	22.08	3.509	20.47	3.363	20.76	3.655
2. Chronically poor if poor for two or more periods	Not Poor	89.47	1.608	89.47	1.608	89.47	1.608	89.47	1.608
(Structural Income)	Poor	4.678	4.240	4.678	4.240	4.678	4.240	4.678	4.240
3. Chronically poor if poor for three or more periods	Not Poor	92.84	2.632	92.84	2.632	92.84	2.632	92.84	2.632
(Structural Income)	Poor	1.316	3.216	1.316	3.216	1.316	3.216	1.316	3.216
4. Chronically poor if poor for four periods	Not Poor	93.42	3.655	93.42	3.655	93.42	3.655	93.42	3.655
(Structural Income)	Poor	0.731	2.193	0.731	2.193	0.731	2.193	0.731	2.193

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Row 1 shows the monetary poverty headcount proposed by Foster, Greer, Thorbecke (1984) for the observed income (Y_{it}). 2/ Rows 2-4 shows the chronic poverty measurement proposed by Foster (2009) assuming that households are considered chronically poor, if they fall in poverty in different number of periods using the structural income (Y_{it}^{S}).

6.2 Results of the analysis of the transitory income Y_{it}^T and the persistence of income shocks

This subsection presents the results of the method proposed by Arellano, Blundell and Bonhomme (2017) for analysing Y_{it}^T. This analysis is an important complement to the analysis in the previous section as here I can understand the persistence of the transitory income, Y_{it}^T, along the whole distribution of the observed income, Y_{it} and conditional to the convergence clubs. This allows me to understand whether the persistence and direction of the income shocks reinforce the dynamics of the structural income, or conversely, they act as opposite forces on the observed income.

In this section I am using the ordinary residuals[12] of the gamma model to estimate the transitory income, Y_{it}^T, in order to maintain the additivity among the structural income and transitory income:

$$Y_{it}^T = \hat{\varepsilon}_{it} = Y_{it} - E\Big\{f\Big(\sum_j \hat{\beta}_j A_{ijt}\Big)\Big\} = Y_{it} - e^{\hat{\beta}_0 + \sum_j \hat{\beta}_j A_{ijt}}$$

Intuitively Arellano, Blundell and Bonhomme (2017) propose as a first step standardizing the income shocks (Y_{it}^T) in the traditional way:

$$Y_{it}^{T*} = \frac{Y_{it}^T - \overline{Y_{it}^T}}{SD_{Y_{it}^T}} \quad (Eq.14).$$

Where $\overline{Y_{it}^T}$ is the sample average of Y_{it}^T over i and t and $SD_{Y_{it}^T}$ is the standard deviation over i and t.

The next step proposed by Arellano, Blundell and Bonhomme (2017) is to run a regression of the standardized income shocks on its own lags in the form of a third degree Hermite polynomial for different conditional quantiles of Y_{it}^{T*} on Y_{it-1}^{T*}. The cth conditional quantile function is defined as:

[12]The ordinary residuals are not recommended as a measure of fitness of a Generalized Linear Model. Instead, McCullagh and Nelders (1989) propose the Pearson residuals, Anscombe residuals or deviance residuals.

$$Q_{Y_{it}^{T*}|Y_{it-1}^{T*}}(c) = \psi_c + \phi_{1,c}Y_{it-1}^{T*} + \phi_{2,c}[(Y_{it-1}^{T*})^2 - 1] + \phi_{3,c}[(Y_{it-1}^{T*})^3 - 3Y_{it-1}^{T*}]$$

$$(Eq.15).$$

where c is the cth quantile of the shock of the conditional quantile function ω_{it} and the parameters to estimate are ψ_c, $\phi_{1,c}$, $\phi_{2,c}$ and $\phi_{3,c}$. In this case, such parameters were estimated for 20 quantiles, i.e $c = 1, 2, ..., 19$. The maximization problem that the conditional quantile regression has to solve in this case is:

$$\hat{\Phi}_c = \arg\min_{\Phi \in \mathbb{R}^4} \sum_{i=1}^{N} \sum_{t=2008}^{2010} \theta_c \{Y_{it}^{T*} - \psi_c - \phi_{1,c}Y_{it-1}^{T*} - \phi_{2,c}[(Y_{it-1}^{T*})^2 - 1] - \phi_{3,c}[(Y_{it-1}^{T*})^3 - 3Y_{it-1}^{T*}]\}$$

$$(Eq.16).$$

where $\hat{\Phi}_c = [\hat{\psi_c}\ \hat{\phi_{1,c}}\ \hat{\phi_{2,c}}\ \hat{\phi_{3,c}}]'$ and θ_c is the tilted absolute value function[13].

Let $Y_{it,r}^{T*}$ be the rth quantile of the entire distribution of the standardized transitory income over all individuals and all years. Arellano, Blundell and Bonhomme (2017) measure the persistence of the income shocks or transitory income Y_{it}^{T} as a 19x19 matrix with element (r, c) given by the derivative of the conditional quantile function $Q_{Y_{it}^{T*}|Y_{it-1}^{T*}}(c)$ with respect to Y_{it-1}^{T*}, and evaluating it on the different r quantiles of the distribution of standardized transitory income, given the non-linear nature of such derivative:

$$\frac{\delta Q_{Y_{it,r}^{T*}|Y_{it-1,r}^{T*}}(c)}{\delta Y_{it-1,r}^{T*}} = \hat{\phi_{1,c}} + 2\hat{\phi_{2,c}}Y_{it-1,r}^{T*} + 3\hat{\phi_{3,c}}[(Y_{it,r}^{T*})^2 - 1] \quad \text{for} \quad r, c = 1, 2, ...19$$

$$(Eq.17).$$

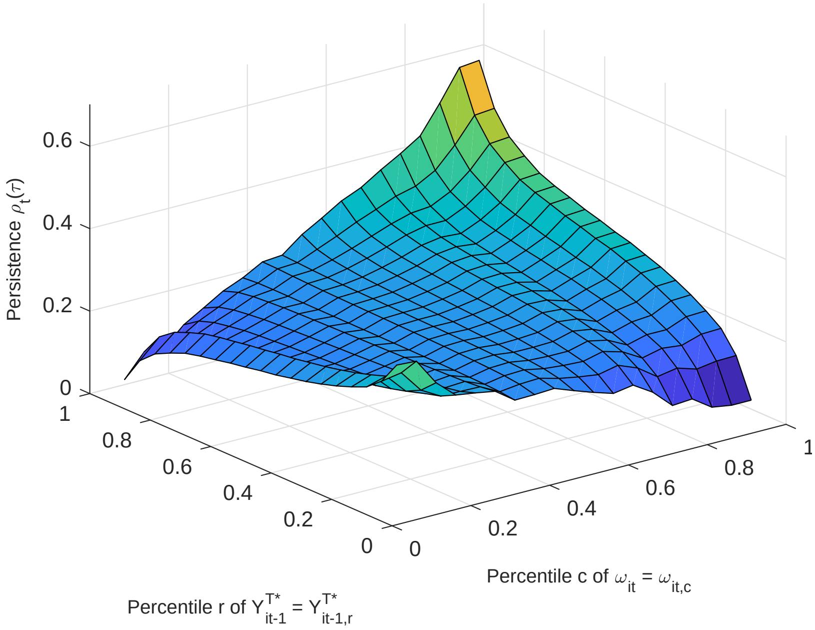
[13] The tilted absolute value function is defined as:

$$\theta_c(\zeta) = \begin{cases} (\tau - 1)\zeta & \text{if} \quad \zeta < 0 \\ \tau\zeta & \text{if} \quad \zeta \geq 0 \end{cases} \quad \text{for } \zeta \in (-\infty, \infty)$$

This equation corresponds to $\rho_t(\tau)$ in 12 and can be interpreted as the average persistence across η when is hit by a current shock ω_{it} with rank c (Arellano, Blundell and Bonhomme, 2017). In other words, if $\omega_{it,c}$ represents the cth quantile of the shock of the conditional quantile function, $\rho_t(\tau)$ is the persistence of Y_{it-1} when it is hit by the shock on the cth quantile, $\omega_{it,c}$.

Figures 5 and 6 shows the estimation of the persistence $\rho_t(\tau)$ as defined by equation 12. That is to say, how the conditional quantile function of current transitory income Y_i^{T*} given Y_{it-1}^{T*} changes, when Y_{it-1}^{T*} changes (Arellano, Blundell and Bonhomme, 2017). The left horizontal axis shows the percentile of Y_{it-1}^{T*} (i.e $Y_{it-1,r}^{T*}$), while the right horizontal axis draws the percentile of the innovation of the quantile process, ω_{it} (i.e $\omega_{it,c}$) after re-scaling the axis between 0 and 1 (instead of the original 0 to 20 due to the 20 quantiles). In line with the findings of Arellano, Blundell and Bonhomme (2017), the graph shows various characteristics: first, the persistence of the conditional quantile autoregressive process changes accordingly to the quantile r of the lagged transitory income ($Y_{it-1,r}^{T*}$) and the quantile c of ω_{it} ($\omega_{it,c}$). Second, the persistence of transitory income (or income shocks) is highest when high transitory income households (i.e high r of Y_{it-1}^{T*} or equivalently positive transitory income) are hit by a positive shock (high c of ω_{it}), and when low-transitory income households (low r of Y_{it-1}^{T*} or equivalently negative transitory income) are hit by a negative shock (low c of ω_{it}) (Arellano, Blundell and Bonhomme, 2017). The first case is the most persistent with a persistence close to 0.6, while in the second case the persistence is around 0.35. Third, negative shocks hitting high transitory income households, and positive shocks hitting low transitory income households are associated with a lower persistence of transitory income, reaching levels of 0.05 (Arellano, Blundell and Bonhomme, 2017).

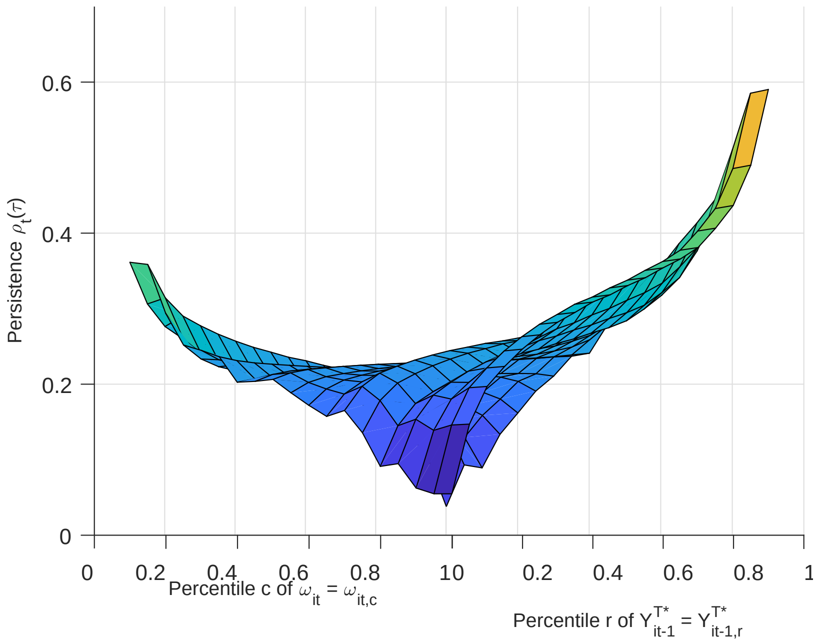
Figure 5: Persistence of a quantile autoregression of transitory income Y_{it}^{T}



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1./ Quantile autoregression of the transitory income Y_{it}^{T*} using the method proposed by Arellano, Blundell and Bonhomme (2017) 2./ Right axis: Percentiles of the shock of the quantile autoregressive process, ω_{it}. Left axis: Percentiles of the lagged standardized transitory income Y_{it-1}^{T*}. Vertical axis: persistence of the transitory income process, $\rho_t(\tau) = \frac{\delta Q_{Y_{it,r}^{T*}|Y_{it-1,r}^{T*}}(c)}{\delta Y_{it-1,r}^{T*}}$.

Figure 6: Persistence of a quantile autoregression of transitory income Y_{it}^{T}

Lateral profile



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1./ Quantile autoregression of the transitory income Y_{it}^{T*} using the method proposed by Arellano, Blundell and Bonhomme (2017) 2./ Right axis: Percentiles of the shock of the quantile autoregressive process, ω_{it}. Left axis: Percentiles of the lagged standardized transitory income Y_{it-1}^{T*}. Vertical axis: persistence of the transitory income process, $\rho_t(\tau) = \frac{\delta Q_{Y_{it,r}^{T*}|Y_{it-1,r}^{T*}}(c)}{\delta Y_{it-1,r}^{T*}}$.

This pattern of persistence on the transitory income coincides with the pattern found by Arellano, Blundell and Bonhomme (2017) for the period 1999-2009 using US data and for the period 2005-2006 with Norwegian data.

Despite these promising results, the analysis in this section is only partial, as I am using the distribution of the transitory income over the whole sample, without taking into consideration how the transitory income is related to the structural income. The possible relation between the structural income and the transitory income is relevant, as, on the one hand, the persistence of the transitory income Y_{it-1}^T may lead to a permanent jump in the observed household per capita income Y_{it}, causing that a household trapped in dynamic structural poverty never appears as being in monetary poverty (according to FGT(0) on the observed income Y_{it}) due to the persistence of such a shock. On the other hand, it may happen that households trapped in dynamic structural poverty also have more persistent negative transitory income pushing them to an even lower observed household per capita income.

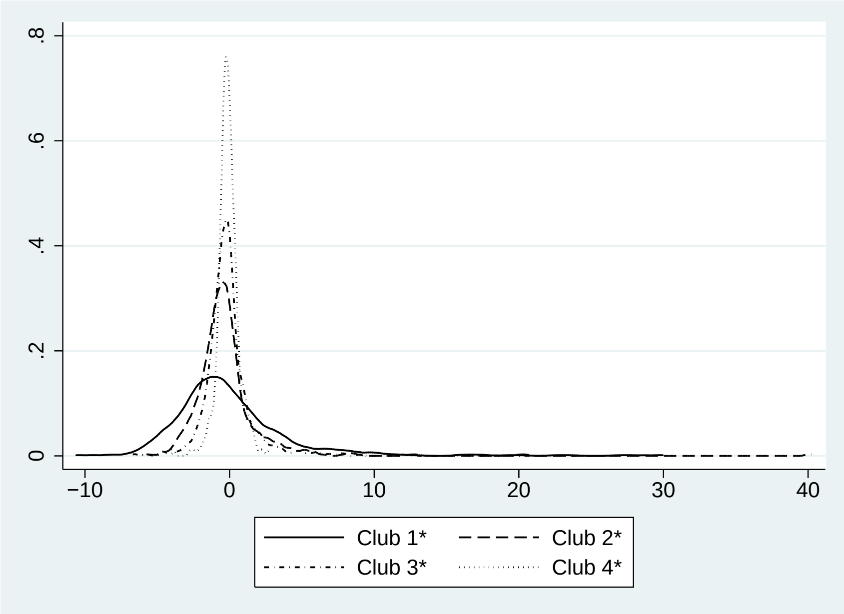
A separate analysis of the transitory and structural income is incomplete as it only presents a partial picture of the poverty status of the households. The next subsection presents a joint analysis of both types of income and identifies patterns of the transitory income within and between convergence clubs.

6.3 Joint analysis of the structural income Y_{it}^S and transitory income Y_{it}^T

The way I propose to analyse jointly the transitory income (Y_{it}^T) and the structural income (Y_{it}^S) is repeating the analysis of the transitory income, conditioning on the convergence club.

Figure 7 shows the density of the transitory income conditional on the merged convergence *club* to which the household belongs to, i.e Y_{it}^T for $i \in G_{M^*}$ and $M^* = 1^*, 2^*, 3^*, 4^*$ where G_{M^*} refers to the merged convergence club. Table 5 presents some basic location and dispersion statistics of these variables. From the figure, it is apparent that the transitory income of each of the four clubs has the same general shape of the distribution as they are centered around the mean with a heavy upper tail. Nevertheless, the mean of the distribution of the transitory income of the households belonging to *club one** is higher than the mean of the distribution of the other three *clubs*. In turn, *club three**, has a highest mean of the distribution compared with the other two clubs, followed by *club four** and *club two**. In other words, the average transitory income of the households trapped in dynamic structural poverty is higher than the average transitory income of the households belonging to *club two** but lower than the average transitory income of the households belonging to *club one**, the club that converges to the highest structural income. Furthermore, the quantiles 10 to 50 of the transitory income of the households belonging to convergence *club four** are higher when compared with the same quantiles of the transitory income for the households belonging to the other three convergence clubs. Additionally, the dispersion of the transitory income is lowest for the households belonging to the *club four** and increases progressively up to reach its highest on the *club one**. Moreover, the median value of the transitory income is negative for the four convergence clubs, implying that households are more prone to face negative income shocks across the entire income distribution.

Figure 7: Conditional Kernel density estimates of transitory household income (Y_{it}^{T}) by merged convergence club of structural income (Y_{it}^{S}) ($Y_{it}^{T} \mid i \in G_{M^*}$ for $M^* = 1^*, 2^*, 3^*, 4^*$)



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Kernel density estimates of the transitory household income (Y_{it}^{T}) using epanechnikov kernel and data driven bandwidth selection method. 2/ The observations of the transitory income (Y_{it}^{T}) are pooled trough time, as every households belongs to only one convergence club across all periods.

Table 5: Descriptive statistics of transitory household income (Y_{it}^{T}) by convergence club of structural income (Y_{it}^{S}) ($Y_{it}^{T} \mid i \in G_{M^*}$ for $M^* = 1^*, 2^*, 3^*, 4^*$))

Club	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)	Mean	S.D	N
Club 1*	-3.768	-2.246	-0.741	1.345	4.323	0.111	4.477	324
Club 2*	-2.250	-1.333	-0.518	0.268	1.865	-0.146	3.121	344
Club 3*	-1.607	-0.885	-0.272	0.343	1.476	-0.0509	2.020	1908
Club 4*	-0.808	-0.475	-0.157	0.221	0.632	-0.125	0.769	160

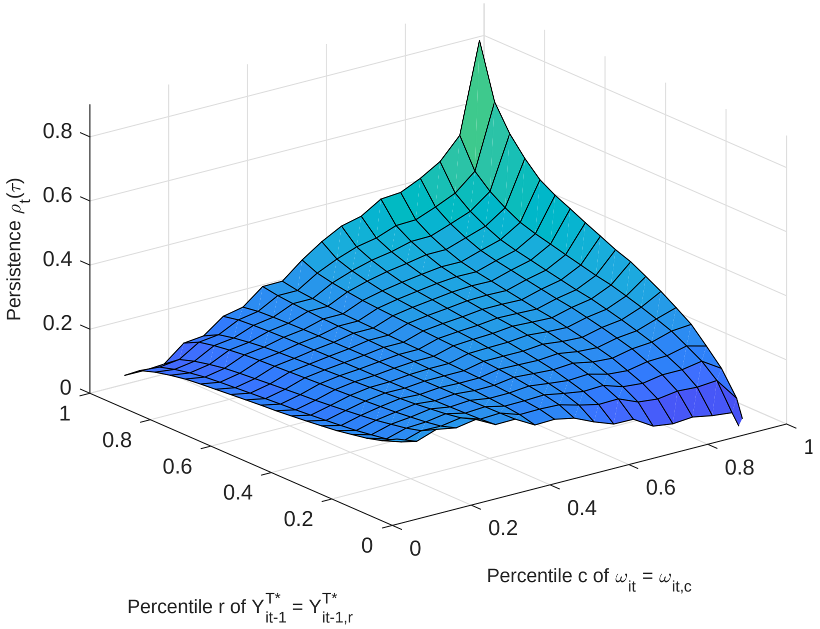
Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ Q(0.10) to Q(0.90) represent the quantile 10 to quantile 90 of the transitory income (Y_{it}^{T}) for each merged convergence club of the structural income (Y_{it}^{S}). 2/ The observations of the transitory income (Y_{it}^{T}) are pooled trough time, as every households belongs to only one convergence club across all periods.

Finally, it is also important to take into account that 58.3% of the households belonging to convergence *club one** have a negative transitory income ($Y_{it}^T < 0$), while this percentage is 66.5% and 62.2% for the households belonging to convergence *clubs two** and *three** respectively. In respect to the convergence *clubs* $four^*$, 61.2% of the households trapped in dynamic structural poverty, have also experienced a negative transitory income.

Appendix 3 and figures 8 and 9 shows the transitory income persistence estimated by convergence club. Clubs *one**, *two** and *three** have a similar persistence pattern than the whole sample. Nevertheless, the persistence of the transitory income of households belonging to convergence *club* $four^*$ (figure 9), follows a similar pattern than that of the whole sample only when households with high transitory income are hit by a positive shock (high c of ω_{it}): the persistence of transitory income is highest (around 0.9) when high-transitory income households (i.e high r of Y_{it-1}^{T*}) are hit by a positive shock (high c of ω_{it}). Nevertheless, the persistence is the lowest when negative transitory income households are hit by a negative shock (between -.0.5 and 0.1). Those are good news for the households belonging to this club, as it implies that the negative occurrences of the transitory income and negative shocks to it, tend to disappear quickly, or even reverse. For the households with positive transitory income and good shocks, the persistence of these, will help the household to leave monetary poverty transiently, for a relatively long period due to its high persistence, but in the long run the dynamics of the structural income will drag them below the poverty line again.

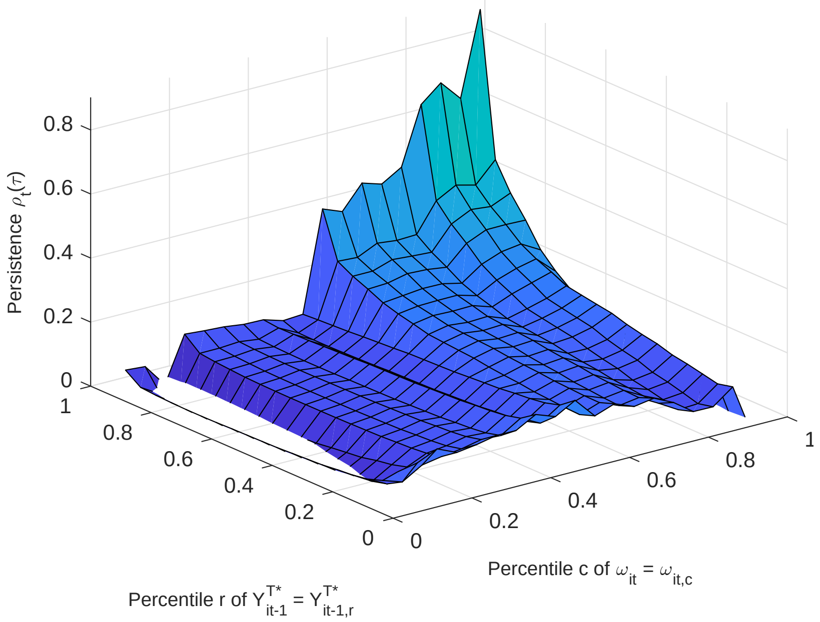
Finally, figure 10 compares the persistence of the transitory income between convergence clubs. The highest persistence of the transitory income is in general for the households belonging to the convergence *club one** and *club two** (compared with the other two convergence clubs). Convergence *club* $four^*$ (households trapped in dynamic structural poverty) presents the lowest persistence of the transitory income, except when they face a very positive transitory income and a very good shock, in which case the transitory income helps to lift them out of monetary poverty transiently, even if the dynamics of the structural income tend to drag them into poverty again.

Figure 8: Persistence of a quantile autoregression of the transitory income for convergence *Club* 3^*
$(Y_{it}^T \mid i \in G_{3^*})$



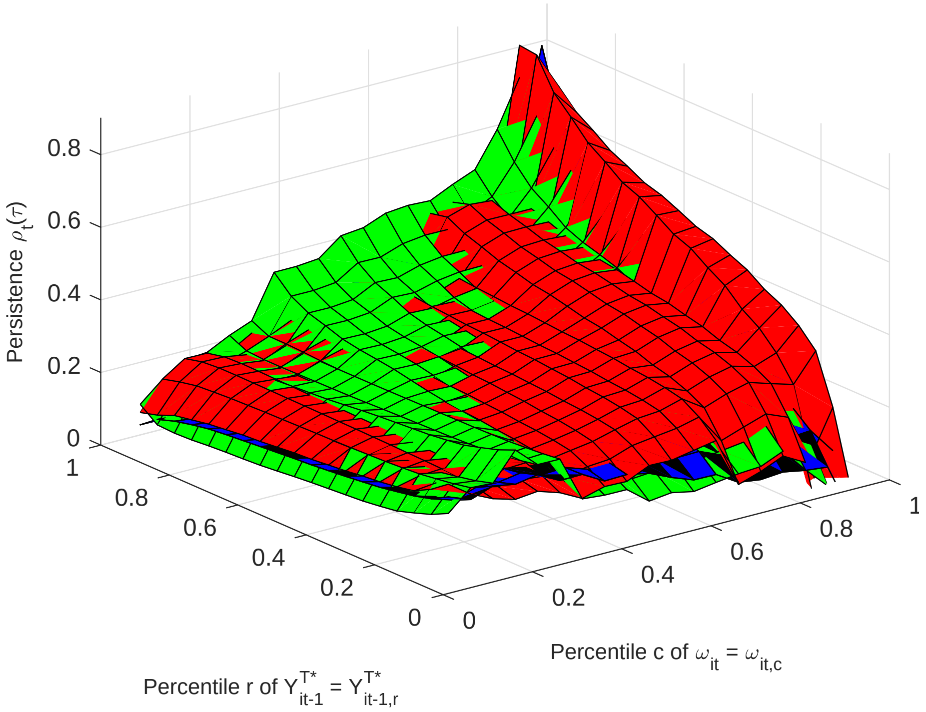
Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1./ Quantile autoregression of the transitory income Y_{it}^{T*} using the method proposed by Arellano, Blundell and Bonhomme (2017) 2./ Right axis: Percentiles of the shock of the quantile autoregressive process, ω_{it}. Left axis: Percentiles of the lagged standardized transitory income Y_{it-1}^{T*}. Vertical axis: persistence of the transitory income process, $\rho_t(\tau) = \frac{\delta Q_{Y_{it,r}^{T*} \mid Y_{it-1,r}^{T*}}(c)}{\delta Y_{it-1,r}^{T*}}$.

Figure 9: Persistence of a quantile autoregression of the transitory income for convergence *Club* 4^*
$(Y_{it}^T \mid i \in G_{4^*})$



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1./ Quantile autoregression of the transitory income Y_{it}^{T*} using the method proposed by Arellano, Blundell and Bonhomme (2017) 2./ Right axis: Percentiles of the shock of the quantile autoregressive process, ω_{it}. Left axis: Percentiles of the lagged standardized transitory income Y_{it-1}^{T*}. Vertical axis: persistence of the transitory income process, $\rho_t(\tau) = \frac{\delta Q_{Y_{it,r}^{T*} \mid Y_{it-1,r}^{T*}}(c)}{\delta Y_{it-1,r}^{T*}}$.

Figure 10: Persistence of a quantile autoregression of the transitory income
Overlapped surfaces
$(Y_{it}^{T} \mid i \in G_{M^*}$ for $M^* = 1^*(green),\ 2^*(red),\ 3^*(blue),\ 4^*(black)))$



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1./ Quantile autoregression of the transitory income Y_{it}^{T*} using the method proposed by Arellano, Blundell and Bonhomme (2017) 2./ Right axis: Percentiles of the shock of the quantile autoregressive process, ω_{it}. Left axis: Percentiles of the lagged standardized transitory income Y_{it-1}^{T*}. Vertical axis: persistence of the transitory income process, $\rho_t(\tau) = \frac{\delta Q_{Y_{it,r}^{T*} \mid Y_{it-1,r}^{T*}}(c)}{\delta Y_{it-1,r}^{T*}}$.

7. Conclusions

This paper has proposed a new methodology for measuring poverty, following the theoretical framework of Carter and Barrett (2006). They propose to measure poverty through the dynamics of the assets that allow a household to produce enough income to be consistently above a given monetary poverty line. The estimation and analysis of such asset dynamics has been previously made following a two-step estimation method: in the first step an asset index -or structural income- is estimated. This paper uses a Generalized Linear Model for estimating such asset index -or structural income-. The second step consist on estimating the dynamics of this asset index -or structural income-. As a novelty, this paper proposes to estimate the dynamics of this asset index using the methodology developed by Phillips and Sul (2007) for classifying countries in convergence clubs. Such methodology was designed for a Solow growth model and proposes to classify countries (in our case households) according to a common growth component of the structural income.

This study also recommends to complement the analysis of the asset index -structural income- with an analysis of the transitory income -residuals of the regression for estimating the asset index-. For this purpose I use the panel quantile autoregressive methodology proposed by Arellano, Blundell and Bonhomme (2017). This method allows to estimate the persistence and direction of the transitory income at which households are subjected.

The measure proposed here brings additional information on the poverty status beyond the traditional poverty headcount -FGT(0)- and the multidimensional poverty index, as it takes into account the dynamic evolution and persistence of the structural income. It is based on the ideas of Friedman's permanent income hypothesis, who recognises that the most important determinant of an individual's consumption is the permanent income, instead of the current income, which is in turn a function of the assets possessed by the individual. From an applied perspective this poverty measurement uses ground breaking econometric methods which can be applied to a wide variety of econometric problems. Finally, as all the fourth generation of poverty measures, the one in this work is based on economic modelling.

The technique proposed here was applied to a panel (with 4 cohorts) of Colombian urban households for the years 2007-2010 for the cities of Bogotá, Bucaramanga and Cali. The results suggests the existence of four convergence clubs for the structural income among urban households on this cities. The first convergence club (*club one**) is for the households with a high quantity of assets at their disposal, that allow them to produce an income well above the monetary poverty line. Not only that, but its estimated dynamics imply that the households in this club will converge to a dynamic equilibrium of the structural income well above the monetary poverty line. The *club two** and *club three** includes households whose assets are in a medium-high and medium low range, below the assets possessed by the *club one**, but above the monetary poverty line. These households have the capacity to produce structural income above the monetary poverty line but below the structural income of the *club one**. It is important to say that the households with the lowest structural income among *club three** may fall temporarily in monetary poverty, as their structural income may fall temporarily below the monetary poverty line, but the estimated growth component will lift them out of poverty.

The final club ($club\ four^*$), is composed by households that do not have enough assets for generating a structural income above the monetary poverty line. Furthermore, the structural income of the households belonging to this club have a dynamic of the structural income that converges below the monetary poverty line. These households may have a structural income temporarily above the monetary poverty line; nevertheless, the (negative) growth pattern of the structural income of this club will pull them below the monetary poverty line on the long run. The households belonging to convergence $club\ four^*$ are said to be trapped in dynamic structural poverty, representing 5.8% of the households in the sample (out of 684 households). This percentage is the new poverty measurement proposed in this paper.

When analysing the transitory income for the full sample (i.e including the households belonging to the four convergence clubs) it is found that there exist differences in the impact of an innovation to the quantile process according to the direction and magnitude of the percentile of the past level of the transitory income. Persistence of transitory income is highest when high-transitory income households (households with positive transitory income) are hit by a good shock, and when low-transitory income households (households with negative transitory income) are hit by a bad shock. In contrast, bad shocks hitting high transitory income households, and good shocks hitting low transitory income households are associated with much lower persistence of transitory income histories.

If the transitory income analysis is repeated separately for each convergence club, the results show that, for convergence *club one**, *two** and *three** (the clubs with the structural income converging above the monetary poverty line) the transitory income have a similar persistence and shock pattern than for the whole sample. For convergence *club four** (the households trapped in structural dynamic poverty) the persistence of transitory income is also highest when high-transitory income households are hit by a good shock. Nevertheless, the persistence is the lowest when negative transitory income households are hit by a bad shock. This implies that negative occurrences of the transitory income and negative shocks to it, tend to disappear quickly, or even reverse. For the households with positive transitory income and good shocks, the persistence of these, will help the household to leave monetary poverty transiently, but in the long run the dynamics of the structural income will drag them below the poverty line again.

Finally, some disadvantages of the method proposed here are its computational complexity, its high requirement of information, as it needs a panel data with at least 4 cohorts of information on household income (or expenditure), a set of demographic controls and its variation through time. Additionally, Phillips and Sul (2007, 2009) method was developed for large samples across the time dimension. Here, the approach for small samples proposed by the authors was used. Nevertheless, a panel data with a longer time dimension is required to draw final conclusions on the performance of the method. Despite this, the method seems to perform well in comparison with other traditional poverty measurements given the behaviour of the dynamics of the structural income of the households between and within convergence clubs.

8.

References

[1] Adato, Michelle, Michael Carter and Julian May. Exploring poverty traps and social exclusion in South Africa using qualitative and quantitative data. *The Journal of Development Studies.* 42(2): 226-247. 2006

[2] Alkire, Sabina and James Foster. Counting and multidimensional poverty measurement. *Journal of Public Economics.* 95(7): 476–487. 2011

[3] Antman, Francisca and David McKenzie. Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. *The Journal of Development Studies.* 43(6): 1057-1083. 2007

[4] Arellano, Manuel, Richard Blundell and Stéphane Bonhomme. Earning and consumpsion dynamics: A nonlinear panel data framework. *Econometrica.* 85(3): 693-734. 2017

[5] Arunachalama, Raj and Ajay Shenoyb. Poverty traps, convergence, and the dynamics of household income. *Journal of Development Economics.* Volume 126: 215-230. 2017

[6] Azarides, Costas and John Stachurski. Poverty traps. In Philippe Aghion and Steven Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1A. 2005

[7] Banerjee, Abhijit and Andrew Newman. Occupational Choice and the Process of Development. *Journal of Political Economy*, 101(2): 274–298. 1993

[8] Barrett, Christopher, Paswel Marenya, John Mcpeak , Bart Minten ,Festus Murithi , Willis Oluoch-Kosura, Frank Place , Jean Claude Randrianarisoa , Jhon Rasambainarivo and Justine Wangila. Welfare dynamics in rural Kenya and Madagascar. *The Journal of Development Studies.* 42(2): 248-277. 2006

[9] Bound, John, Charles Brown and Nancy Mathiowetz. Measurement Error in Survey Data. *Handbook of Econometrics.* 5: 3705-3843. 2001

[10] Bourguignon, Francois and Christian Morrison. Inequality among World Citizens: 1820-1992. *American Economic Review.* 92(4):727–744. 2002

[11] Carter, Michael and Christopher Barrett. The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies*. 42(2): 178-199. 2006

[12] Dasgupta, Partha and Debraj Ray. Inequality as a Determinant of Malnutrition and Unemployment: Theory. *The Economic Journal*. 96(384): 1011–1034. 1986

[13] Duncan, Greg, and Willard Rodgers. Has Children's Poverty Become More Persistent?. *American Sociological Review*. 56: 538-50. 1991

[14] Dutta, Swati. Identifying Single or Multiple Poverty Trap: An Application to Indian Household Panel Data. *Social Indicators Research*. 120(1): 157–179. 2015

[15] FEDESARROLLO. Ficha técnica: Encuesta Social Longitudinal de Fedesarrollo 2010. Fundación para la Educación Superior y el Desarrollo, Bogotá, Colombia. 2010

[16] Foster, James. A Class of Chronic Poverty Measures, in *Poverty Dynamics: Interdisciplinary Perspectives*, ed. by Addison, Tony, David Hulme and Ravi Kanbur, 59–76. Oxford: Oxford University Press. 2009

[17] Foster, James, Joel Greer, and Erik Thorbecke. A Class of Decomposable Poverty Measures. *Econometrica*. 52(3): 761-66. 1984

[18] Galor, Oded and Joseph Zeira. Income Distribution and Macroeconomics. *Review of Economic Studies*. 60(1): 35–52. 1993

[19] Giesbert, Lena and Kati Schindler. Assets, Shocks, and Poverty Traps in Rural Mozambique. *World Development*, 40(8): 1594-1609. 2012

[20] Gomez, Javier and Alejandro Lopez. Poverty dynamics in rural Mexico: What does the future hold?. *Ensayos Revista de Economía*. 32(2): 55-74. 2013

[21] Hien, Vu. Income Distribution and Poverty Traps in Vietnam. Mimeo. CIFREM, University of Trento. 2011

[22] Jalan, Jyotsna and Martin Ravallion. Geographic poverty traps? A micro model of consumption growth in rural China. *Journal of Applied Econometrics*. 17(4): 329-346. 2002

[23] Koenker, Roger and Zhijie Xiao. Quantile Autoregression. *Journal of the American Statistical Association*. 101(475): 980-990. 2006

[24] Krugman, Paul. Increasing Returns and Economic Geography. *Journal of Political Economy*. Volume 99: 483 – 499. 1991

[25] Kwak, Sungil and Stephen Smith. Multidimensional poverty and interlocking poverty traps: Framework and application to ethiopian household panel data. Institute for International Economic Policy Working Paper Series, 2011(4). 2011

[26] Kwak, Sungil and Stephen Smith. Regional Agricultural Endowments and Shifts of Poverty Trap Equilibria: Evidence from Ethiopian Panel Data. *The Journal of Development Studies*. 49(7): 955-975. 2013

[27] Levy, Frank. How Big is the American Underclass?. Working Paper 0090–1, Urban Institute, Washington, DC, USA. 1977

[28] Lybbert, Travis, Christopher Barrett, Solomon Desta and Layne Coppock. Stochastic wealth dynamics and risk management among a poor population. *The Economic Journal*. 114: 750–777. 2004

[29] Matsuyama, Kiminori. Financial Market Globalization, Symmetry-Breaking, and Endogenous Inequality of Nations. *Econometrica*. Volume 72: 853–884. 2004

[30] Maxwell, Daniel, Bapu Vaitla, Girmay Tesfay and Nigussie Abadi. Resilience, Food Security Dynamics, and Poverty Traps in Northern Ethiopia Analysis of a Biannual Panel Dataset, 2011–2013. Working paper, Feinstein International Center. 2013

[31] McCullagh, Peter and John Nelder. *Generalized Linear Models*. Chapman & Hall/CRC: Monographs on Statistics and Applied Probability. 1989

[32] Mookherjee, Dilip and Debraj Ray. Contractual Structure and Wealth Accumulation. *American Economic Review*. 92(4):818-849. 2002

[33] Mukasa, Adamon. Food Price Shocks-induced Poverty Traps: Analysis Using a Panel Dataset From Uganda. Presented in Commodity market instability and asymmetries in developing countries: Development impacts and policies. Clermont-Ferrand, France. 2015

[34] Muñoz, Manuel. Estimación de equivalencias de escala. *Revista de Economía Institucional*. 16(30): 99-130. 2014

[35] Murphy, Kevin, Andrei Shleifer and Robert Vishny. Industrialization and the Big Push. *Journal of Political Economy*. 97(5): 1003 – 1026. 1989

[36] Naschold, Felix. The Poor Stay Poor: Household Asset Poverty Traps in Rural Semi-Arid India. *World Development*, 40(10): 2033-2043. 2012

[37] Panopoulou, Ekaterini and Theologos Pantelidis. Club Convergence in Carbon Dioxide Emissions. *Journal of Political Economy*. 44(1):47-70. 2009

[38] Phillips, Peter. Challenges of trending time series in econometrics. *Mathematics and Computers in Simulation*. 68: 401-416. 2005

[39] Phillips, Peter and Donggyu Sul. Transition modeling and econometric convergence tests. *Econometrica*. 75(6): 1771-1855. 2007

[40] Phillips, Peter and Donggyu Sul. Economic transition and growth. *Journal of applied econometrics*. 24: 1153-1185. 2009

[41] Quah, Danny. Convergence Empirics across Economies with (Some) Capital Mobility. *Journal of Economic Growth*.Volume 1: 95 – 124. 1996

[42] Quisumbing, Arnes and Bob Baulch. Assets and Poverty Traps in Rural Bangladesh. *The Journal of Development Studies*. 49(7): 898-916. 2013

[43] Rodriguez, Jorge and Claudio Gonzalez. Household Income Dynamics and Poverty Traps in El Salvador. Presented in AAEA Annual Meetings 2004. The Ohio State University.

[44] Sandoval, Carlos. Welfare and poverty traps in urban Colombia: an approach from the asset dynamics perspective. Working paper. 2019

[45] You, Jung. Risk, under-investment in agricultural assets and dynamic asset poverty in rural China *China Economic Review*, 29: 27-45. 2014

[46] Zerfu, Daniel. Poverty Traps and Institutions in Ethiopia. Working paper. African Development Bank. 2012

[47] Zhou, Li and Calum Turvey. Testing Asset Dynamics for Poverty Traps in Rural China. *Canadian Journal of Agricultural Economics*. 63(2015): 129–162. 2015

9. Appendix

9.1 Appendix 1

Fedesarrollo Longitudinal Social Survey
Estimation of the Household Per capita Income Adjusted by Adult Equivalence and Divided by the Monetary Poverty Line (l_{it})

Generalized Linear Model
Distribution: Gamma. Link Function: Logarithmic. 2007-2010.
Main Urban Areas

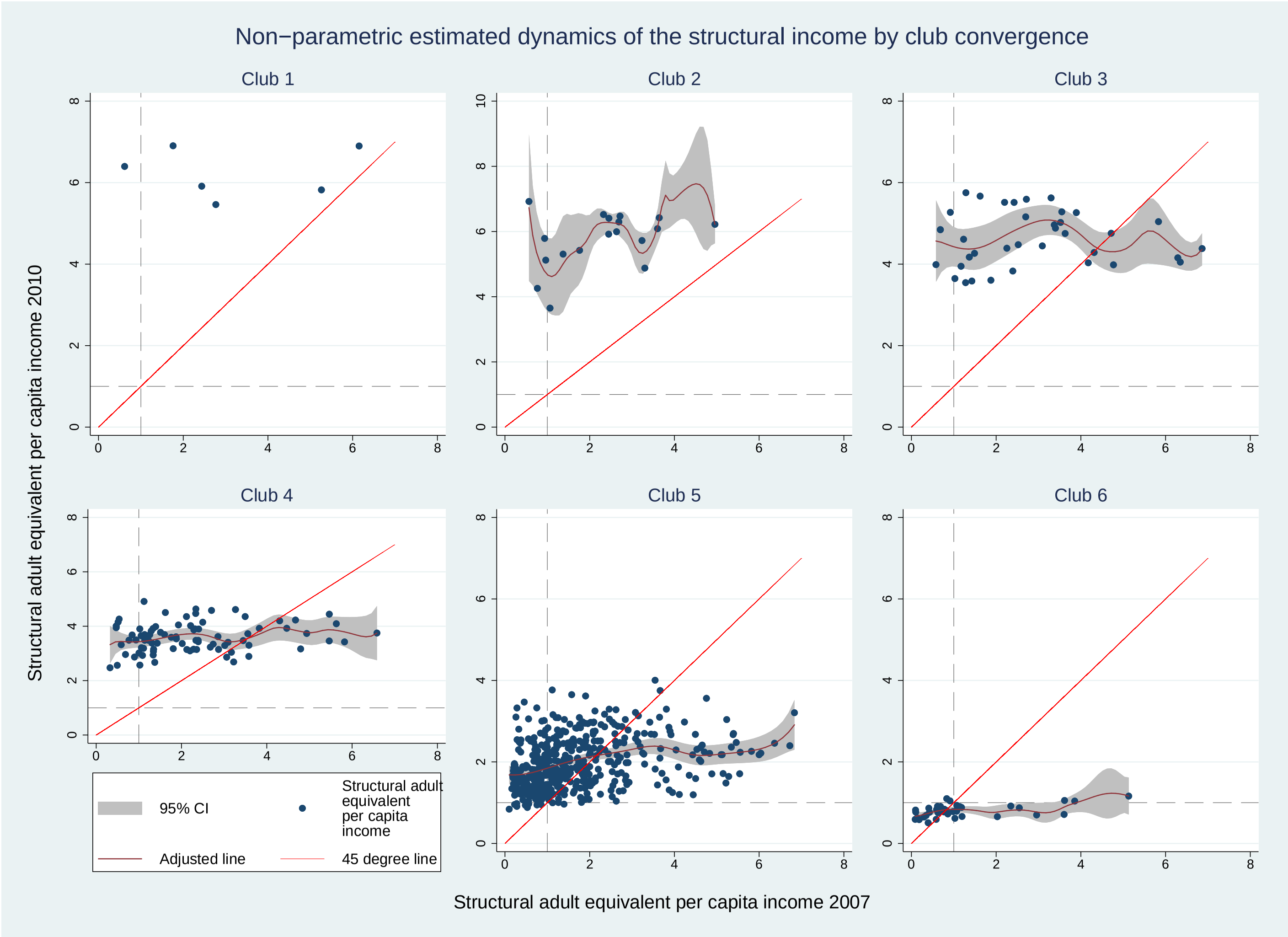
	(1)	(2)	(3)	(4)
VARIABLES	GLM (2007)	GLM (2008)	2009	GLM (2010)
Male head of the household	0.2615***	0.0241	0.1413**	0.0966
	(0.080)	(0.074)	(0.069)	(0.081)
Years of Education of head of the household	0.0810***	0.0946***	0.0763***	0.0822***
	(0.008)	(0.007)	(0.007)	(0.008)
Age of head of the household	0.0180	0.0203	-0.0012	0.0154
	(0.018)	(0.017)	(0.017)	(0.019)
Age of head of the household 2	-0.0000	-0.0000	0.0001	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Employment rate among household members	2.3668***	0.8849***	0.8285***	0.8985***
	(0.172)	(0.132)	(0.126)	(0.145)
Proportion of members between 0 and 12 years old	-0.8003***	-0.8006***	-1.2901***	-1.1849***
	(0.230)	(0.219)	(0.209)	(0.258)
Proportion of members 62+ years old	-0.7842***	0.0774	0.0475	0.2000
	(0.183)	(0.167)	(0.155)	(0.173)
Number of self-employed household members	0.1437***	-0.0073	0.0073	0.0553
	(0.043)	(0.041)	(0.039)	(0.044)
Bogota	0.0423	0.0982	0.0844	0.1883
	(0.184)	(0.171)	(0.161)	(0.189)
Cali	-0.0008	0.2006	0.1195	0.1531
	(0.185)	(0.172)	(0.162)	(0.190)
Constant	-2.6859***	-1.7823***	-0.8590*	-1.5529***
	(0.542)	(0.516)	(0.504)	(0.582)
Observations	684	684	684	684
F	.	.	.	.

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1/ The table presents the first step in the two-step approach proposed by Adato, Carter and May (2006). 2/ Columns 1-4 present the estimated coefficients of the control variables using a Generalized Linear Model with a Gamma distribution and a Logarithmic link function. 3/ Estimations for all the years available (2007, 2008, 2009, 2010). 4/ For the precise definition of the variables used here see section 4 (data section). 5/ A dummy variable for the city of Bucaramanga is left out

9.2 Appendix 2

Figure 11: Club convergence result on the Structural (or predicted) income (Y_{it}^S) 2007 vs 2010



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey (FLSS) for the years 2007 (horizontal axe) and 2010 (vertical axe). Note: 1/ The graph presents the second step in the multi-step approach proposed by Adato, Carter and May (2006) using the convergence club classification method by Phillips and Sul (2007). The vertical axis variable is the structural income (Y_{it}^S) in 2010 and the horizontal axis variable is the structural income (Y_{it}^S) in 2007. 2/ The adjusted model is a nonparametric model estimated using a kernel-weighted local polynomial smoothing regression with a rectangle kernel, a 4th degree approach polynomial and the bandwidth selected by a rule of thumb for each club separately. 3/ The dashed vertical and horizontal line represent 1, i.e, the point where the structural income (Y_{it}^S) is equal to the monetary poverty line. 4/ The Phillips and Sul (2007) algorithm found 6 convergence clubs.

9.3 Appendix 3

Figure 12: Persistence of a quantile autoregression of the transitory income for convergence *Club* 1^*
$(Y_{it}^T \mid i \in G_{1^*})$

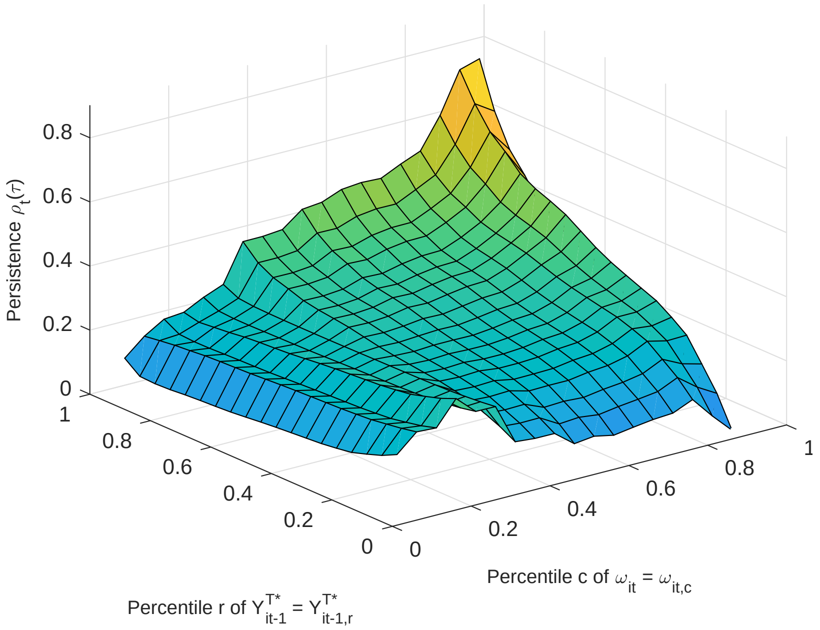
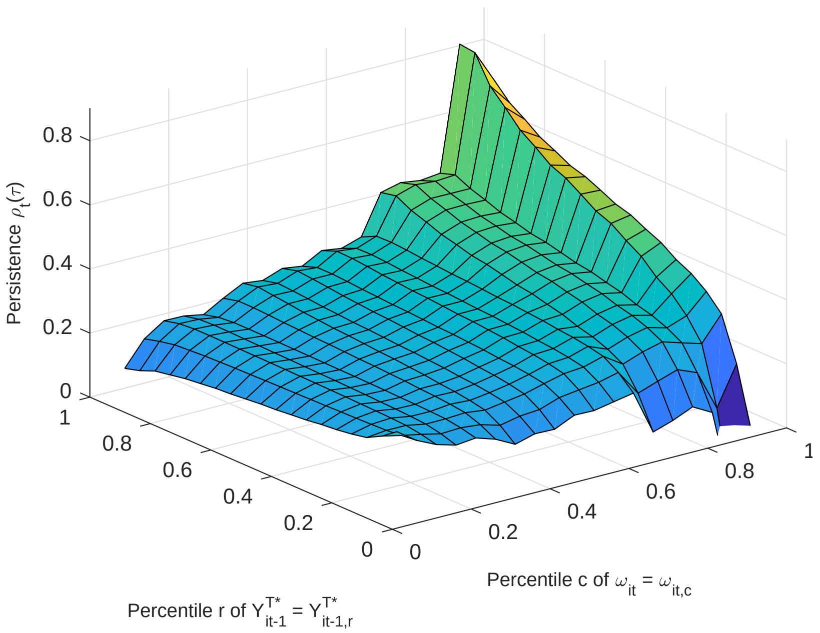


Figure 13: Persistence of a quantile autoregression of the transitory income for convergence *Club* 2^*
$(Y_{it}^T \mid i \in G_{2^*})$



Source: Author's estimations using Fedesarrollo Longitudinal Social Survey. Note: 1./ Quantile autoregression of the transitory income Y_{it}^{T*} using the method proposed by Arellano, Blundell and Bonhomme (2017) 2./ Right axis: Percentiles of the shock of the quantile autoregressive process, ω_{it}. Left axis: Percentiles of the lagged standardized transitory income Y_{it-1}^{T*}. Vertical axis: persistence of the transitory income process, $\rho_t(\tau) = \frac{\delta Q_{Y^{T*}_{it,r}|Y^{T*}_{it-1,r}}(c)}{\delta Y^{T*}_{it-1,r}}$.

Chapter 4. The effect of different types of health insurance on health outcomes, medical care use, and risk protection: evidence from Colombia

Abstract

There are three types of health insurance status for the Colombian population: individuals with private insurance, individuals on public funded health insurance for the poor and the uninsured population. Previous studies have found that right after its introduction the health insurance for the poor increased the use of preventive healthcare services when comparing them with individuals belonging to the private insurance, or those who were uninsured. This paper re evaluates the differences in the effect of the public health insurance for the poor, the private health insurance and the uninsured population in Colombia along a wide variety of dimensions, using recent data from the Colombian Longitudinal Survey by Universidad de los Andes (CLSA) for the years 2010, 2013 and 2016 through a Fuzzy Regression Discontinuity design. The results show that in 2010 and onwards, there is no significant difference among the three possible health insurance types in outcomes related to health status, medical care use, risk protection against illness and behavioural distortions in urban areas. This may be due to the universal coverage reached by the health system in 2009 and the equalization of the scope among the three health insurance types proposed by the public health policy at the end of the 2000's.

Keywords: Health insurance, risk protection, regression discontinuity.

JEL classification: I12, I13, I18, C14, C21, C26

1. Introduction

Access to affordable healthcare is one of the main components of any egalitarian society, as it protect its members from negative financial shocks due to catastrophic illnesses. Most of the countries belonging to the OECD have reached universal health coverage which is also fully public funded in many cases (World Health Organization and World Bank, 2017). Nowadays the idea of universal health coverage have expanded to developing countries were public funding is scarce and healthcare can not be fully funded by the government. Nevertheless, intermediate solutions to the funding problem have been worked out. There are systems were the part of the population with enough income pays its own private health insurance, while the poorest part of the population receives public funded or subsidized healthcare. Colombia is one of those cases.

In 1993 the Colombian health system went through a deep reform with the goal of reaching universal healthcare access, in particular for the most vulnerable population. Prior to the reform, only 25% of the population had access to formal health insurance (Miller et. al, 2013), while for 2009 universal health coverage was reached, mainly through the expansion of a health insurance scheme for the poorest population (Ministerio de Salud y Protección Social, 2012), called the Subsidized Health Regime (SR). The Subsidized Health Regime is financed through taxes and cross subsidies from individuals who can afford paying the health insurance scheme from private companies, i.e, individuals on the Contributive Health Regime (CR).

The initial goal of the government was to cover the poorest part of the population with health insurance and, after reaching universal coverage for the poor, equalize the Subsidized and Contributive Health Regime in terms of quality, use of medical services, financial risk protection and health status of the population belonging to each regime. In Escobar et. al (2009) words, "according to the law, the supply-side subsidies should gradually transform into demand-side subsidies as insurance coverage expanded, eventually leading to universal coverage with a uniform package for everyone". In 2009, the universal healthcare coverage was reached and afterwards the main goal of the public health policy has been to equalize the reach of the rights, medicines and services for which the users are entitled in both, the Contributive and Subsidized Health Regimes.

Previous studies have shown that before reaching universal health coverage, those affiliated to the Subsidized Health Regime had a greater medical care use (Trujillo, Portillo, and Vernon, 2005), better self-reported health status, more preventive and curative outpatient care, and fewer hospitalizations (Gaviria, Medina and Mejía, 2007), access to prenatal and neonatal care services (Giedion et al, 2009), a higher birthweight (Camacho and Conover, 2013) and a lower financial risk (Miller et al., 2013).

In the evaluation of the Subsidized Health Regime, all of the studies mentioned above use as control group a combination of the individuals belonging to the Contributive Health Regime and Uninsured individuals, leading to some difficulties: when using such combination of individuals as control group the effect on some variables may go in one direction due to one subgroup, while it may go on the other direction due to the other subgroup.

Additionally, the most recent evaluation using Regression discontinuity methods was made using data from 2003 and 2005 (Miller et al, 2013). At that moment, the subsidized regime was still expanding and didn't cover the complete eligible population. Furthermore, a positive result on the evaluation of the Subsidized Health Regime close to the time of its implementation may be due to the improvement in the health care services for those who were previously uninsured, without clarifying what was the effect in respect to those who were affiliated to the Contributive Health Regime.

The present paper re evaluates the effect of the Subsidized Health Regime in Colombia, recognizing the fact that there may be differences among the control groups when evaluating the Subsidized Health Regime. In other words, this paper evaluates the different effect of the three insurance status, i.e belonging to the Subsidized Health Regime, Contributive Health Regime or being uninsured, on a wide selection of variables following the Miller et. al (2013) categorization, including proxys for health status (i.e circulatory system diseases, other chronic diseases excluding those related with the circulatory system, infectious diseases, activities of daily living and health events during last year), behavioural distortions (i.e fruit consumption. vegetable consumption, fried food consumption and physical activity), medical care use variables and risk protection and consumption smoothing (i.e total, health and food expenditure). The difference in the effect is tested using a Fuzzy Regression Discontinuity (FRDD) design.

Indeed, given the design of the Subsidized Health Regime this is the appropriate econometric technique to to find the effect of each health insurance type on different response variables. In particular, the assignation to the Subsidized Health Regime is done through a poverty-targeting index called SISBEN score (Miller et.al, 2013). If a household has a score of the SISBEN index below a certain threshold it is eligible to the Subsidized Health Regime in a non deterministic way[1], creating the perfect setting for using the FRDD method. The SISBEN score can operate as an instrument for the affiliation status to each type of health insurance. After having said estimation on the first stage, a second stage estimation may be conducted using the prediction of the affiliation status as exogenous variable on a Fuzzy Regression Discontinuity that uses as endogenous variables a set of response variables including health status, behavioural distortion, medical care use and risk protection related variables.

Additionaly, this paper uses more recent data (compared with other studies) from the Colombian Longitudinal Survey by Universidad de los Andes (CLSA) for the years 2010, 2013 and 2016 -after reaching universal health coverage in Colombia in 2009-. This implies that the results should focus more on finding whether the different health insurance regimes have been equalized according to the policy goal of the government.

The results support the health insurance regime equalization' hypothesis and allow me to conclude that the differences that existed among the Subsidized Health Regime and its control group, do not exist any more, after reaching universal healthcare coverage. Furthermore, there are no significant differences among being affiliated to the Contributive Health Regime, Subsidized Health Regime or being uninsured. This is due mainly to the fact that after reaching universal healthcare, the healthcare providers should provide basic health services, even to the population that is not insured.

Various robustness checks were conducted to test the stability of the results.

[1] Not everybody with a score below the eligibility threshold of the poverty-index is on the Subsidized Health regime and some households are part of it irrespectively of their score for administrative rules that will be explained in section 2.2.

First, different specifications for the approximation degree, bandwidth selection and kernel function used on the Fuzzy Regression Discontinuity setting were attempted. Second, the models were estimated for each year separately, and only for the years 2010 and 2013. Third, the models were estimated for the head of the household and partner only. Fourth, a Bonferroni test were conducted for taking into account the sequentiality of the test carried for each one of the three health insurance regimes. Fifth, the models were estimated only for insured individuals, dropping the uninsured population. The results are always the same: there is no difference on the response variables according to what type of health insurance the individual possesses.

The remainder of this paper is organized as follows: section 2 presents the policy background, section 3 makes a brief literature review of previous studies, section 4 describes the data; section 5 describes the econometric method and the results of the evaluation of the three possible health insurance types in Colombia. Finally, section 6 presents briefly the conclusions and outlook.

2. Policy background

The modern Colombian health system had its origins in 1993, when the prior health system went through a structural reform in the context of a decentralization and other state modernization reforms (Escobar et. al, 2009).

Prior to the Reform in 1993, roughly 25% of Colombians (a proportion of the population with formal jobs) had some kind of "explicit health insurance" (Miller et. al, 2013). The remaining 75% of the Colombian population had what Miller et. al (2013) call "implicit insurance" as they could effectively use medical care services in public institutions paying only a part of the cost-recovering fee and given that out-of-pocket payments were based on economic status. Nevertheless, more than half of total spending on health was out of pocket and used to represent an actual barrier to access medical care services, specially among the poorest population (Escobar et. al, 2009). The health care system was funded by tax revenue, payroll contributions, and out-of-pocket expenditures with no pooling of the funding or the risk (Escobar et. al, 2009).

Under Law 100 in 1993, three types of health insurance affiliation were created: first, the Contributive Regime (CR), aimed at formal employees and their families, or independent workers with enough income to pay for health insurance. It is financed with the premium from companies and workers. Premiums are tied to payroll in the case of formally employed individuals, or income for the case of independent workers. In both cases they should pay 12.5% of their income as premium. It is mandatory for anyone formally employed (Ministerio de Salud y Protección Social, 2012; Camacho and Conover, 2013). Second, the Subsidized Regime (SR) for the poor and unemployed with no ability to pay. Funded by a cross subsidy from workers in the formal sector and contributions of the central and local governments through taxes. The Subsidized Regime works through subsidies at the demand. The selection of the population eligible for the Subsidized Regime is made through the SISBEN survey (Beneficiary Identification System) and with census of vulnerable individuals like those belonging to indigenous and displaced population (Ministerio de Salud y Protección Social, 2012; Camacho and Conover, 2013). These individuals do not make any insurance contributions (do not pay premium). Third, the Special Health Regime (SpecR) for workers of some public institutions such as the army, the national oil company (Ecopetrol), teachers in public schools, employees of public universities, the police force and some other public servants (Ministerio de Salud y Protección Social, 2012). These three types of health insurance affiliation provide health services and medicine plans. Lastly, the uninsured individuals are known as vinculados and "can go to hospitals and clinics for public assistance" (Camacho and Conover, 2013).

As the system is designed to subsidise the demand, the insured individuals may choose the insurer, care provider and receive a set of health benefits called the Mandatory Health Plan (MHP)[2] (Escobar et. al, 2009). Insured individuals can enrol their family unit. Escobar et. al (2009) estimate that for the contributory regime the health insurance package had a premium of U$207 annually in 2007 covering all lev-

[2]The Mandatory Health Plan is comprised by the rights and services that a person affiliated to either the contributive or subsidized health regime is entitled. It covers the family of the affiliated member of the household including the spouse or permanent partner of the affiliated; those children under 18 years of either the spouse or the affiliated who are part of the household and are economically dependent from the couple; children over 18 years with permanent disability or those with less than 25 years that are full-time students and financially dependent on the affiliated. If there is no spouse, permanent partner or entitled children, the health insurance family coverage may be extended to parents of the affiliated that have no pension and depend on the affiliated economically (Ministerio de Salud y Protección Social, 2012).

els of care, while the premium was U$117 for the subsidized regime covering "primary care, some inpatient care, and emergency care". After reaching the near-universal coverage of the system, the idea was to close the gap in the MHP between the Subsidized and Contributive health regime.

Miller et. al (2013) describe the SR "as a variant of the 'managed competition' model". Enthoven (1993) summarizes such model as one where an intermediary (the government in this case), acting on behalf of the users of the healthcare system, regulate the market to promote price competition among insurance companies. The government "establishes rules of equity", selects the mandatory health plan (MHP), is in charge of enrolling users of the healthcare into the system, define the respective price-elastic demand, and "manages risk selection". The organizations under this model should integrate financing functions (represented by the insurance companies here) and the delivery of healthcare services (hospitals, clinics and doctors hired by the insurance company).

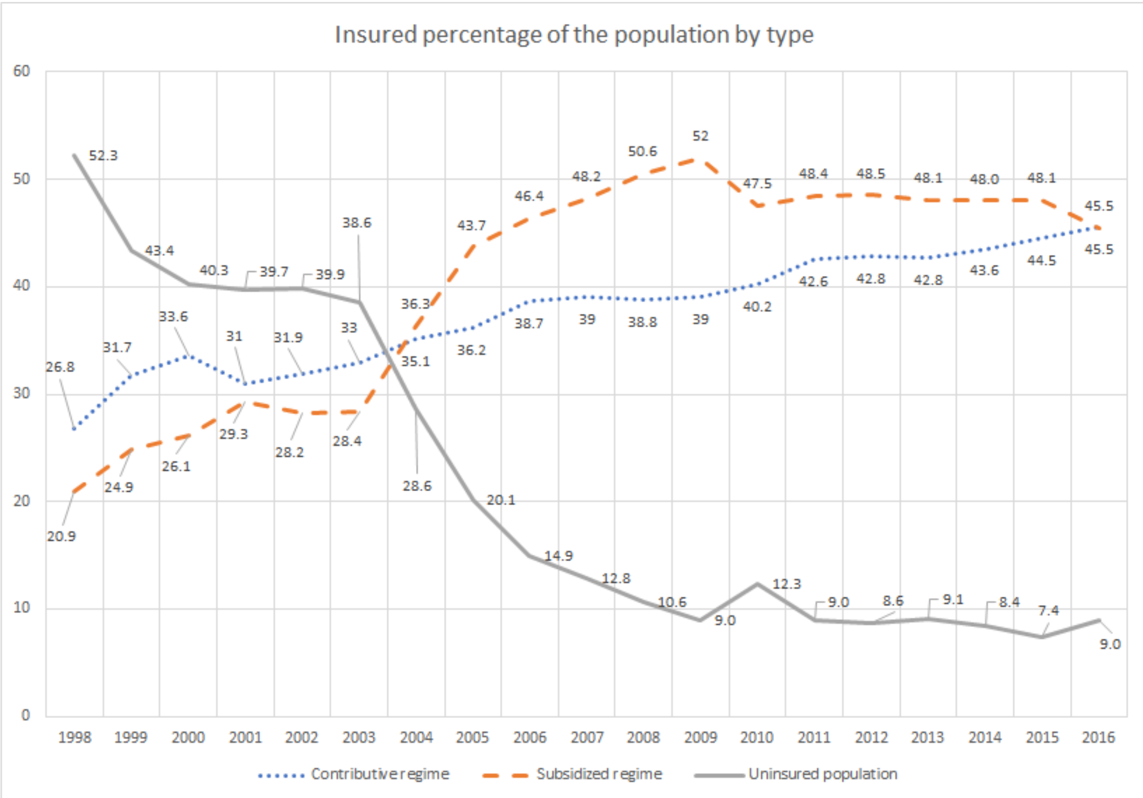
The health insurance coverage jump in Colombia from around 25% of the population in 1993 to 91% in 2009, mainly due to the rapid growth of the Subsidized Health Regime (Miller et. al, 2013). Thereafter the general insurance coverage has fluctuated around 90% of the total population, showing mainly re-compositions among SR and CR, but holding the total percentage of insured population approximately constant (see figure 1).

Once the near-universal coverage of the formal health insurance system in Colombia were reached in 2009 the policy focused on equalizing the scope of the MHP among the Contributive and Subsidized Health Insurance Regime. The MHP for the Contributive Regime and the Subsidized Regime were different in reach when the Subsidized Regime was first implemented due to budgetary limitations. As expected, the MHP used to cover a wider range of procedures and medicines for the Contributive Regime (Ministerio de Salud y Protección Social, 2012).

The equalization process of the MHPs started as an implicit rule of investment for the resources initially intended to the expansion of the SR once the universal coverage was reached. It started in 2009 with the population of 18 years old and younger and in 2012 it ended when the rest of the population was included (Ministerio de Salud y Protección Social, 2017).

To assure the permanent equalization of the MHPs among both health insurance regimes such equalization was introduced as a permanent law in 2011 (Ministerio de Salud y Protección Social, 2012).

Figure 1: Insured Population



Source: Ministerio de Salud y Protección Social (2012).

2.1 Price structure: Premium, copayment, sliding scale fees, deductibles and cost-recovery fees

In the Colombian Health System the population and beneficiaries are subject to premium, copayments, sliding scale fees, cost-recovery fees and deductibles according to what regime the individuals are affiliated (or if they are insured or not at all).

For the Contributive Regime the payment changes according to weather the person is the directly affiliated to the Health System (the contributing working member) or one of the beneficiaries (part of the family of the affiliated that can be covered by the health insurance of the affiliated). The premium is paid by all of those who are formally employed and by the independent workers with a monthly income higher than the minimum legal monthly wage. The formally employed should contribute 12.5% of the total wage, while the independent workers should contribute 5% (or 12.5% of 40%) of the total monthly income.

There are two types of fees that should be paid by all the Health System customers: the copayment and sliding scale fees. The later is applied with the sole objective of rationalizing the use of system (Ministerio de Salud y Protección Social, 2012). The copayment, which is the value paid by the individual every time he uses a health service, should be paid only by the beneficiary members (not by the affiliated member) depending on the income range of the affiliated member [3].The sliding scale fee, should be paid by both, the affiliated and the beneficiaries, per each event [4]. For the beneficiaries, those payments help financing of the Mandatory Health Plan. In any case the sliding scale fees nor the copayments can be a barrier to access the health services for the poorest population. For this reason the people belonging to the level 1 of the SISBEN (SISBEN ranges between level 1 and 6, being 1 for the poorest people) in the Subsidized Health Regime are excluded from the moderator fees and copayments (Ministerio de Salud y Protección Social, 2012) [5]. The uninsured people have to pay a cost-recovery fee every time they need medical attention [6].

[3]If the affiliated's member income ranges between 1 and 2 minimum legal wages, the beneficiary should pay 11.5% of a minimum legal daily wage (MLDW), if the income goes from 2 to 5 minimum legal wages the copayment should be 17.3% of a MLDW. Finally, if the wage is higher than 5 minimum legal wages the copayment should be 23% of a MLDW. The copayments amount has a maximum limit for each event and per year (Ministerio de Salud y Protección Social, 2004)

[4]If the affiliated's member income ranges between 1 and 2 minimum legal wages, the individual should pay 11.7% of a MLDW, if the income goes from 2 to 5 minimum legal wages the sliding scale fee should be 46.1% of a MLDW. Finally, if the wage is higher than 5 minimum legal wages the moderation fee should be 121.5% of a MLDW (Ministerio de Salud y Protección Social, 2004)

[5]Nevertheless, people belonging to the SISBEN level 2 and 3 have copayments of 10% of the total cost of the service with a maximum amount of 50% of a minimum legal monthly wage per event and a maximum cost of one minimum legal monthly wage per year (Ministerio de Salud y Protección Social, 2004).

[6]Population non-affiliated to the SR belonging to the SISBEN level 1 would pay 5% of the total cost without exceeding one minimum legal monthly wage per event. People belonging to the SISBEN level 2 would pay 10% of the total cost without exceeding two minimum legal monthly wage per event. Individuals belonging to the SISBEN level 3 would pay up to 30% of the total cost without exceeding three minimum legal monthly wage per event. People belonging to the SISBEN levels 4, 5 or 6 or with payment capacity would pay the full cost of the healthcare service per event

Copayments are applied to all services covered by the MHP in the Contributive Regime, with the exception of health promotion and prevention services, maternal and child care, care of communicable diseases, catastrophic or high cost illness and emergency care (Ministerio de Salud y Protección Social, 2012). Any service, activity, procedure and intervention included in the Subsidized Mandatory Health Plan is free and there are no moderator fees or copayment for the children during the first year of life, for everyone of SISBEN level 1 (at any age) and special populations [7]. Attention is also free in the Subsidized Regime for services including the prenatal care, delivery care, preventive health services, control programs in care of communicable diseases, catastrophic diseases and dental consultation (Ministerio de Salud y Protección Social, 2012).

2.2 Eligibility for the Subsidized Regime

As previously explained the eligibility for the SR depends on whether a household is above or below a preset threshold in a poverty-targeting index called SISBEN [8] (Miller et al., 2013). The SISBEN phase I (1995-2005) was based on fourteen components capturing different aspects of socio-economic status at the household level [9] (Department of National Planning, 2008a). A SISBEN score per household was then calculated by using principal components across this fourteen dimensions [10]. The corresponding SISBEN score of the households were then clustered into 6 levels, and those households belonging to the levels 1 and 2 were eligible for the SR (Miller et al., 2013) (See Figure 2).

Nevertheless, local politicians and potential beneficiaries had incentives to manipulate the SISBEN score for increasing the enrolment to the SR [11] (Camacho and

(Ministerio de Salud y Protección Social, 2013).

[7]Special populations include indigenous communities, workers of the public kindergartens, displaced population and homeless population.

[8]SISBEN is the acronym in Spanish for Sistema de Identificación de Beneficiarios

[9]Dwelling material, access to public services, assets possession, demographic composition, education, and labour market variables (Miller et al., 2013).

[10]The scale of the scores goes from 0 to 100 where the wealth of the household increases as the score approaches 100 (Miller et al., 2013)

[11]In 1998 the algorithm to calculate the SISBEN score was released to the officials in charge of making the surveys for measuring eligibility at county level. At the same time, the distribution of the SISBEN score phase I started showing a sharp discontinuity on the SISBEN score at the cutoff

Conover, 2011) which, combined with the low discriminatory power of some variables made necessary to reform the SISBEN index. With this in mind, the SISBEN phase II (2005-2011) was introduced. In this version of the SISBEN, the principal components methodology and the thresholds for the full eligibility to the SR remained unmodified (Department of National Planning, 2008a) while the set of variables used for calculating the score changed. Between 2009 and 2011 partial health subsidies were introduced to the strata 3 of the SISBEN phase II [12]. After 2011 the SISBEN phase III was implemented for improving the discrimination power of the index. This version of the index is estimated using fuzzy sets and it focuses in the quality of life and the capacities of the individuals to generate welfare (Department of National Planning, 2008a). There are now three domains: 14 main cities, other municipalities and rural areas [13] (See Figure 2).

Miller et. al (2013) argue that the correlation between eligibility for the SR and effective enrolment is not perfect due to errors in household classification, manipulation of SISBEN scores, lack of public funding, enrolment prior to the SISBEN survey baseline (Miller et. al, 2013) and automatic inclusion to the SR of special populations irrespectively of their SISBEN score. For the SISBEN II and SISBEN III, there was still some manipulation of the index at a lesser degree than the SISBEN I index. From 2006 the budget of local governments was not an obstacle anymore for the enrolment of eligible individuals and since 2009 Colombia has reached universal coverage on healthcare services (Ministerio de Salud y Protección Social, 2012) [14]. This had implications for the identification strategies on past works on the topic (Miller et. al, 2013), and will have implications on the identification strategy followed on this paper, as we will be able to assume that every person who is eligible for the SR has a chance to be selected for the SR greater than 0.

thresholds.

[12]This implies that households with SISBEN scores between 48 to 58 in the urban areas and 30 to 45 in the rural areas started to have access to partial subsidies in the SR (see Figure 2).

[13]Households are eligible for the SR if they are below 54.86 on the 14 main cities, 51.57 on other municipalities and 37.80 on rural areas.

[14]Adding up the number of insured individuals among CR, SR and SpecR.

Figure 2: Elegibility to the Subsidized Regime according to SISBEN score

Phase	Used to determine elegibility to SR ...	Minor changes	Area	Level	Score
SISBEN phase I	from 1995 to 2003-2005		Urban	1	0 to 36
				2	36 to 47
			Rural	1	0 to 18
				2	18 to 30
SISBEN phase II	from 2005 to 2011	2005 to 2009	Urban	1	0 to 36
				2	36 to 47
			Rural	1	0 to 18
				2	18 to 30
		2009 to 2011 (Includes partial subsidies for SISBEN level 3	Urban	1	0 to 36
				2	36 to 47
				3	47 to 58
			Rural	1	0 to 18
				2	18 to 30
				3	30 to 45
SISBEN phase III	from 2011 to ????		14 main cities	1	0 to 47.99
				2	48 to 54.86
			Other municipalities	1	0 to 44.79
				2	48 to 51.57
			Rural	1	0 to 32.98
				2	32.99 to 37.80

Source: Department of National Planning (2008a,2008b), Ministerio de Salud y Protección Social (2004).

3 Literature review

Previous empirical papers have studied Colombia's SR, but very few works have compared explicitly the SR, CR and the uninsured individuals. In particular three works use propensity score matching as their main econometric method. First, Trujillo, Portillo, and Vernon (2005) compare SR enrollees to uninsured individuals, finding that the SR is associated with greater medical care use. Second, Giedion and Uribe (2009) and third, Giedion et. al (2009) estimate the impact of the SR on healthcare access and use. For a wide range of variables the authors found that being affiliated to the SR increases the access to healthcare services and reduces the number of individuals reporting not seeking medical care for financial reasons.

Gaviria, Medina, and Mejía (2007) found that that SR enrolment is associated with better self-reported health, fewer hospitalizations and more medical care use through instrumental variables. Santa María et al. (2015) evaluate Colombia's health reform implemented in 1993 using a diff in diff setting. They found that the introduction and increase in coverage of the SR has a positive effect on access to health services, health outcomes and quality of service.

Three previous works have been done using a regression discontinuity design. Matching official SISBEN score data with birth records, Camacho and Conover (2013) find that SR enrolment is associated with increased birth weight and better APGAR scores but not prenatal care use, medical supervision of deliveries, or probability of hospital delivery. Miller et. al (2013) find that the SR was able to mitigate the financial risk among the poor and increased the use of preventive healthcare services with significant improvements in health status. De la Mata and Gaviria (2015) measure the exogenous effect of losing health insurance coverage when turning 18 years old. They find that the lack of coverage increases the visits to the emergency room and reduces the use of preventive and curative healthcare services.

Finally, Houweling et al. (2015) use Poisson regression to measure the effect of the health insurance coverage scheme (SR, CR or uninsured) of the mother in neonatal mortality in Colombia. As expected, the authors find a lower mortality rate among infants born from mothers in the CR compared with uninsured mothers or mothers in the SR. They conclude that despite the SR not being sufficient to “close the gap in newborn mortality between socioeconomic groups", it is an improvement in respect with those uninsured.

This paper differs from the existing literature and contributes to it in at least two fronts. First, it re evaluates the effect of the Subsidized Health Regime in Colombia, but modelling explicitly the fact that may be differences among the control groups when evaluating it. In particular it evaluates the differences among the three main health insurance regimes that a person can be part of: Contributive Health Regime, Subsidized Health Regime or Uninsured (excluding the SpecR which represent a minority of the population). Second, this paper uses more recent data compared with similar studies. In particular it uses household surveys for the years 2010, 2013 and 2016, after reaching universal coverage of the Health System in Colombia in 2009. This implies that the results should focus more on finding whether the different health insurance regimes have been equalized according to the initial goal of the government. In particular, after 2009 the government has been implementing what is called the unification of the Mandatory Health Plan (MHP) among the users of the Contributive and Subsidized Health Regime.

4. Data

Colombian Longitudinal Survey by Universidad de los Andes

This paper uses the same data than Sandoval (2019). For this reason some of the text here is similar to the data section in that study.

The Colombian Longitudinal Survey by Universidad de los Andes (CLSA) had its first round in 2010, a second round in 2013 and its more recent round in 2016. It is designed as a panel survey where some of the households in the original cohort (2010) are intended to be followed through time. On its first round, it interviewed in total 10.168 households with representativeness on five regions (Atlántica, Pacífica, Central, Oriental y Bogotá), where 53.5% of the households were located in urban areas and 46.5% in rural areas (Universidad de los Andes, 2011; Sandoval, 2019). In 2013 the sample was composed by 9.262 households. Of these households, 8.848 were intended to be part of the panel in 2010 and were able to be effectively interviewed (Universidad de los Andes, 2014). In 2016, 8.818 households were interviewed, 50% on each area. This implies that the attrition rate of the survey was 6% during the period 2010-2013 and 4.8% between 2013-2016 (Universidad de los Andes, 2017).

This work uses these three cohorts to evaluate the effect of the different health insurance regimes in Colombia, using the pooled sample, i.e, treating each observation as independent across time. This gives a potential sample of 28.248 observations corresponding to each household on each year. Moreover, most of the response variables evaluated here are asked at individual level, which implies that I have a potential of 118.824 observations at individual level. Nevertheless due to the sample design of the survey not all variables are asked to all members of the household, and not all variables are asked on all the cohorts. Basic data like gender and age is asked to all household members in all cohorts, but detailed data like labour status, economic status and health insurance information is asked only to certain members of the household [15].

Given this sampling design, and considering also the non-response rates in the survey, it is expected to have a smaller sample in respect to the numbers described in the previous paragraph. Appendixes 1A-1F shows the exact definition, target population and cohort availability for each of the questions that were used to construct the variables used in the econometric models. Table 1 shows the descriptives of the variables used for estimating the econometric models. Panels A to D of table 1 shows the descriptive statistics of the response variables of the econometric section, divided according to Miller et. al (2013) in health status variables, medical care use variables, behavioural distortion variables and risk protection variables.

Of the 28.248 household-year observations, 53% are urban, comprising the final sample of interest in the econometric models due to reasons that will be explained

[15] In 2010 such detailed information is asked only to the head of the household and partner regarding themselves and children 0 to 9 years old. In 2013, said information was collected for the head of the household and partner in 2013 (that may be different from the persons with the same status in 2010), children from 0 to 13 years old that were selected as part of the panel in 2010 and people from 0 to 64 years old that were not selected as part of the panel in 2010. For 2016, this information set is asked to head of the household and partner in 2010 selected as part of the panel, children 6 to 16 years old that were part of the panel, people from 0 to 64 years old that were not selected as part of the panel in 2010 and persons 10 to 16 years old that were selected as part of the panel.

later. Unfortunately, not all those households have the required variables for the econometric models I am running. This fact, combined with the changing structure of the survey and the non response rates, leave me with a final effective sample of 629 observations, for the variable with less observations and 5.089 for the variables with most observations [16]. In terms of the structure of the survey it asks for welfare conditions, demographic conditions, health and education for the head of the household and its partner in 2010 and for every member of the household in 2013.

Panel A of table 1 shows the descriptive statistics for the health status variables. Such health status variables can also be divided in categories according to circulatory system diseases [17], chronic diseases [18], infectious diseases [19], difficulties to perform activities of daily living [20] and health events during last year [21]. In most of the cases the health status variables are balanced according to the health insurance status, meaning that the incidence of each of those health conditions is similar among the individuals belonging to the CR, SR and uninsured[22]. The only variable that shows imbalance among insurance status are the hypertension which have an incidence of 10% for uninsured people while in the CR is 15% and in the SR is 17%.

Panel B shows descriptives for medical care use variables [23]. Three variables in this category show some type of imbalance among household insurance types. First, 64% of individuals in the SR has visited the doctor during last year, while 70% have

[16] In total, this paper is estimating the effects of the SR, CR and uninsured status on 59 variables. This implies that there is a lot of variation in the number of effective observations we have for each variable. This is due to the fact that each estimation requires information on three variables: The SISBEN score, the insurance affiliation status of the individual and the response variable. The main issue is that not all of the individuals has observations for the three variables. Additional to this not all the variables exist in all years. 27% were asked in all years, and the other 73% only in 2013 and 2016. 50% of the variables have around 1.783 observations, while 15% have 4.084 observations, 15% have 1.547 observations and only one variable has 629 observations. The variable with 629 observations have such a small sample due that it was collected only in 2013 and 2016 at a household level. 58 out of 59 variables have a number of observations between 1.449 and 5.089.

[17] Including thrombosis, hearth attack, hearth conditions and hypertension.

[18] Including asthma, emphysema, diabetes, ulcer, epilepsy and cancer.

[19] Tuberculosis only.

[20] Including movement difficulties, showering difficulties, learning disabilities and blindness.

[21] Including illness, accident, surgery or pregnancy.

[22] To establish an objective criteria I consider that the variables are balanced among different health insurance regimes if the incidence on each regime do not differ by more than 5%.

[23] This category includes a question on whether the person has been hospitalized during the last 12 months, and the use of preventive medicine including visits to doctor, visits to the dentist, visit to the optometrist, use of alternative medicine, use of contraceptives and the use of other preventive medicine.

done it in the contributive regime and 47% among uninsured people. Following the same pattern, 50% of individuals in the SR visited the dentist in the last year, while 55% did it in the CR and only 36% among the uninsured. A pattern in the same direction was followed by those who visited the optometrist.

Panel C of table 1 shows the descriptives for the behavioural distortion variables [24]. Most of the variables in this category do not show differences according to the insurance type. Nevertheless, uninsured individuals consume more fruit in lower frequencies than the individuals in the other two insurance regimes, while the daily fruit consumption is higher for the individuals in the CR. In respect to the fried food, the uninsured individuals consume less fried food twice a week. Finally, the uninsured individuals tend to have less moderate physical activity during the last week compared to either the households in the SR and CR.

Panel D shows the descriptives for risk protection and consumption smoothing. In this case, as expected, the households in the CR have the highest household monthly expenditure, followed by the uninsured households who have the second highest total expenditure per capita, and finalizing with the households in the SR who have the lowest per capita expenditure. In the same direction, the households on the CR have the highest health annual expenditure, followed by the households in the SR, while the uninsured households have the lowest annual health expenditure. The household monthly food expenditure is higher for households in the CR, followed by households in the SR and uninsured households.

[24] This category includes fruit consumption, vegetable consumption, fried food consumption and physical activity

Table 1: Descriptive statistics for all individuals included in econometric models. Urban households (1)

PANEL A : HEALTH STATUS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SUBSIDIZED REGIME			CONTRIBUTIVE REGIME			UNINSURED		
VARIABLES	mean	sd	N	mean	sd	N	mean	sd	N
Thrombosis [2013 and 2016]	0.0129	0.113	1,084	0.0176	0.132	624	0	0	75
Hearth attack [2013 and 2016]	0.0221	0.147	1,084	0.0176	0.132	624	0.0267	0.162	75
Hearth condition [2013 and 2016]	0.0526	0.223	1,084	0.0433	0.204	624	0.0667	0.251	75
Hypertension (Yes) [2013 and 2016]	0.172	0.377	1,084	0.151	0.358	624	0.107	0.311	75
Hypertension (Only during pregnancy) [2013 and 2016]	0.00830	0.0908	1,084	0.00962	0.0977	624	0.0133	0.115	75
Asthma [2013 and 2016]	0.0323	0.177	1,084	0.0288	0.168	624	0.0133	0.115	75
Emphysema [2013 and 2016]	0.0148	0.121	1,084	0.0256	0.158	624	0.0133	0.115	75
Diabetes (Yes) [2013 and 2016]	0.0664	0.249	1,084	0.0497	0.217	624	0.0267	0.162	75
Diabetes (No) [2013 and 2016]	0.928	0.259	1,084	0.944	0.230	624	0.973	0.162	75
Diabetes (Only during pregnancy) [2013 and 2016]	0.00554	0.0742	1,084	0.00641	0.0799	624	0	0	75
Ulcer [2013 and 2016]	0.0793	0.270	1,084	0.0577	0.233	624	0.0667	0.251	75
Epilepsy [2013 and 2016]	0.00738	0.0856	1,084	0.00321	0.0566	624	0	0	75
Cancer [2013 and 2016]	0.0148	0.121	1,084	0.0144	0.119	624	0	0	75
Tuberculosis [2013 and 2016]	0.00369	0.0607	1,084	0.00481	0.0692	624	0	0	75
Movement difficulties [2010, 2013 and 2016]	0.0174	0.131	2,651	0.00696	0.0831	1,294	0.0147	0.121	204
Showering difficulties [2010, 2013 and 2016]	0.00792	0.0887	2,651	0.00309	0.0555	1,294	0.0147	0.121	204
Learning disability [2010, 2013 and 2016]	0.00993	0.0992	2,719	0.00689	0.0827	1,307	0.0189	0.136	212
Blindness [2010, 2013 and 2016]	0.000736	0.0271	2,719	0	0	1,307	0.00472	0.0687	212
Illness [2010, 2013 and 2016]	0.226	0.419	2,550	0.198	0.399	1,295	0.180	0.385	239
Accident [2010, 2013 and 2016]	0.0157	0.124	2,550	0.0247	0.155	1,295	0.00837	0.0913	239
Dentistry [2010, 2013 and 2016]	0.0482	0.214	2,550	0.0502	0.218	1,295	0.0460	0.210	239
Surgery [2010, 2013 and 2016]	0.00588	0.0765	2,550	0.00927	0.0959	1,295	0.00418	0.0647	239
Pregnancy [2013 and 2016]	0.00443	0.0665	902	0.00417	0.0645	480	0	0	67

PANEL B : MEDICAL CARE USE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SUBSIDIZED REGIME			CONTRIBUTIVE REGIME			UNINSURED		
VARIABLES	mean	sd	N	mean	sd	N	mean	sd	N
Hospitalized in the last 12 months [2010, 2013 and 2016]	0.0996	0.300	2,550	0.0849	0.279	1,295	0.0460	0.210	239
Doctor [2010, 2013 and 2016]	0.647	0.478	2,550	0.703	0.457	1,295	0.473	0.500	239
Dentist [2010, 2013 and 2016]	0.508	0.500	2,550	0.553	0.497	1,295	0.360	0.481	239
Optometrist [2010, 2013 and 2016]	0.158	0.365	2,550	0.291	0.454	1,295	0.130	0.337	239
Alternative Medicine [2010, 2013 and 2016]	0.0133	0.115	2,550	0.0247	0.155	1,295	0	0	239
Contraceptive [2010, 2013 and 2016]	0.0905	0.287	1,547	0.0918	0.289	937	0.0855	0.281	152
Other [2013 and 2016]	0.0547	0.227	1,976	0.0549	0.228	875	0	0	117

Source: Author's calculations using CLSA for the years 2013 and 2016 or 2010, 2013 and 2016 when indicated. Note: 1./ Sample divided according to Health Insurance Status 2./ Restricting the sample only to individuals in the econometric models. 3./ Urban households only

Table 1: Descriptive statistics for all individuals included in econometric models. Urban households (2)

PANEL C : BEHAVIORAL DISTORTIONS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SUBSIDIZED REGIME			CONTRIBUTIVE REGIME			UNINSURED		
VARIABLES	mean	sd	N	mean	sd	N	mean	sd	N
Fruit consumption: Less than once a week [2013 and 2016]	0.0775	0.267	1,084	0.0369	0.189	624	0.0933	0.293	75
Fruit consumption: Once a week [2013 and 2016]	0.0996	0.300	1,084	0.0657	0.248	624	0.120	0.327	75
Fruit consumption: 2-4 times a week [2013 and 2016]	0.337	0.473	1,084	0.250	0.433	624	0.293	0.458	75
Fruit consumption: 5-6 times a week [2013 and 2016]	0.0590	0.236	1,084	0.0625	0.242	624	0.0533	0.226	75
Fruit consumption: Daily [2013 and 2016]	0.290	0.454	1,084	0.386	0.487	624	0.347	0.479	75
Fruit consumption: More than once a day [2013 and 2016]	0.137	0.344	1,084	0.199	0.399	624	0.0933	0.293	75
Vegetable consumption: Less than once a week [2013 and 2016]	0.0655	0.248	1,084	0.0417	0.200	624	0.0667	0.251	75
Vegetable consumption: 2-4 times a week [2013 and 2016]	0.368	0.483	1,084	0.346	0.476	624	0.360	0.483	75
Vegetable consumption: 5-6 times a week [2013 and 2016]	0.0655	0.248	1,084	0.0897	0.286	624	0.0933	0.293	75
Vegetable consumption: Daily [2013 and 2016]	0.330	0.471	1,084	0.394	0.489	624	0.333	0.475	75
Vegetable consumption: More than once a day [2013 and 2016]	0.0793	0.270	1,084	0.0833	0.277	624	0.0533	0.226	75
Fried food: Less than once a month [2013 and 2016]	0.00846	0.0916	946	0.0131	0.114	535	0.0152	0.123	66
Fried food: Once a month [2013 and 2016]	0.0317	0.175	946	0.0411	0.199	535	0.0606	0.240	66
Fried food: 2-3 times a month [2013 and 2016]	0.0560	0.230	946	0.0598	0.237	535	0.0909	0.290	66
Fried food: Twice a week [2013 and 2016]	0.263	0.441	946	0.252	0.435	535	0.197	0.401	66
Fried food: 3-4 times a week [2013 and 2016]	0.258	0.438	946	0.267	0.443	535	0.258	0.441	66
Fried food: 5-6 times a week [2013 and 2016]	0.0529	0.224	946	0.0636	0.244	535	0.0606	0.240	66
Fried food: Daily [2013 and 2016]	0.146	0.353	946	0.131	0.338	535	0.121	0.329	66
Fried food: Twice a day [2013 and 2016]	0.0243	0.154	946	0.0224	0.148	535	0.0152	0.123	66
Fried food: Three or more a day [2013 and 2016]	0.00211	0.0460	946	0	0	535	0	0	66
Moderate physical activity: Yes, in the last 7 days [2013 and 2016]	0.0775	0.267	1,084	0.111	0.314	624	0.0933	0.293	75
Moderate physical activity: Number of days, in the last 7 days [2013 and 2016]	0.272	1.127	1,084	0.348	1.247	624	0.213	0.920	75
Moderate physical activity: Minutes per day [2013 and 2016]	5.009	23.89	1,083	7.963	33.66	621	7.703	37.18	74
Strong physical activity: Yes, in the last 7 days [2013 and 2016]	0.0609	0.239	1,084	0.0673	0.251	624	0.0800	0.273	75
Strong physical activity: Number of days, in the last 7 days [2013 and 2016]	0.160	0.793	1,084	0.176	0.854	624	0.227	0.953	75
Strong physical activity: Minutes per day [2013 and 2016]	5.120	24.59	1,081	6.280	29.24	621	5.600	21.95	75

PANEL D : RISK PROTECTION AND CONSUMPTION SMOOTHING

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SUBSIDIZED REGIME			CONTRIBUTIVE REGIME			UNINSURED		
VARIABLES	mean	sd	N	mean	sd	N	mean	sd	N
Household total monthly expenditure [2010, 2013 and 2016]	786.0	530.9	3,180	1,107	695.6	1,620	800.7	494.9	289
Household food monthly expenditure [2010, 2013 and 2016]	397.3	214.6	3,180	465.5	248.6	1,620	388.6	221.8	289
Household health annual expenditure [2013 and 2016]	262.7	564.4	401	499.3	973.4	210	201.5	160.9	18

Source: Author's calculations using CLSA for the years 2013 and 2016 or 2010, 2013 and 2016 when indicated. Note: 1./ Sample divided according to Health Insurance Status 2./ Restricting the sample only to individuals in the econometric models 3./ Urban households only

5 Econometric strategy

Given the design of the health insurance system in Colombia, the appropriate method to evaluate the difference among different insurance types is the Fuzzy Regression Discontinuity design. In this case it is clear that there is a jump in the probability for being affiliated to the Subsidized Health Regime as soon as the household is below a predetermined threshold in the SISBEN score. Nevertheless, the jump in such probability is not from zero to one, given that there are some other rules for inclusion in the subsidized regime that do not depend on the SISBEN score. For instance, on the one hand, if an individual has formal employment he should be affiliated to the Contributive Regime irrespectively of the SISBEN score of the household. On the other hand, indigenous communities and workers of public kindergartens are automatically affiliated to the Subsidized Regime irrespectively of their SISBEN score. Additionally, after 2009 (the implementation of the SISBEN phase II), households belonging to the SISBEN level 3 started being eligible for the Subsidized Health Regime receiving partial subsidies.

As explained in section 3, there may also be administrative reasons for an individual belonging to a particular health insurance type to be misclassified. Those reasons include, among others, classification errors on the Subsidized Health Regime and the lack of funding to enrol all the individuals who where supposed to be enrolled on the Subsidized Health Regime. Those situations were common during the initial phase of the implementation of the SR. Nevertheless, some households may still not be affiliated to the Subsidized Health Regime, despite being eligible (Miller et. al, 2013).

This section is largely based on Imbens and Lemieux (2007), Lee and Lemieux (2010) and Calonico et.al (2014a, 2014b) who explain in detail the ideas behind the Sharp Regression Discontinuity Design and Fuzzy Regression Discontinuity Design, the estimation procedure and inference methods. Most of the equations and exposition here reproduce those in this papers.

Lee and Lemieux (2010) argue that to estimate the treatment effect in this quasi experimental design, it is necessary to combine the Fuzzy Regression Discontinuity (FRD) method with IV. In the IV framework, the estimated treatment effect can be interpreted as the local average treatment. Following Imbens and Lemieux (2007) notation let Y_i be the observed outcome of interest for the individual i (for instance, health status variables). Let $W_i \in \{0, 1\}$ be a variable representing the treatment status, i.e, $W_i = 0$ if unit i was not part of the treatment group, and $W_i = 1$ otherwise (belonging to the SR, CR or been uninsured in this case). In addition, Imbens and Lemieux (2007) define a vector of covariates represented by (X_i, Z_i), where X_i is a scalar and Z_i is a vector with m elements. Neither X_i nor Z_i have been influenced by the treatment status of the individual i. While Z_i can be seen as a set of regular covariates, X_i plays the most important role in the regression discontinuity design, as it is assumed that the assignment to the treatment (in this case the insurance type) is determined, partially or totally by the value of X_i "being on either side of a fixed threshold" c (Imbens and Lemieux, 2007). X_i in this setting corresponds to the SISBEN score. Finally, Lee and Lemieux (2010) propose to define a variable T_i, where $T_i = 1[X_i \geq c]$ if the assignment variable take a value above the eligibility threshold c for the individual i. If that is the case, the probability of receiving the treatment can be written as $Pr(W_i = 1|X_i = x) = \gamma + \delta T_i + g(X_i - c)$ where $g(.)$ is any continuous function in the assignment variable (Lee and Lemieux, 2010).

Lee and Lemieux (2010) define the Fuzzy Regression Discontinuity Design (FRDD) as the equations system:

$$W_i = \gamma + \delta T_i + g(X_i - c) + \nu_i \quad (Eq.1).$$

$$Y_i = \alpha + \tau W_i + f(X_i - c) + \epsilon_i \quad (Eq.2).$$

where γ, δ, α and τ are parameters to be estimated, ν_i and ϵ_i are error terms independent of X while ϵ_i is also independent from W_i. In this setting the most important parameter is τ which can be interpreted as the treatment effect.

In the context of this paper Y_i is one variable in the set of response variables which can be grouped into health status variables, behavioural distortion variables, medical care use variables and risk protection and consumption smoothing variables. For each variable belonging to each group, either $Y_i(1)$ or $Y_i(0)$ is observed depending on whether the individual i belongs to the treatment group ($W_i = 1$) or the control group ($W_i = 0$) (Imbens and Lemieux, 2007). In this particular case, there are three possible treatments: belonging to the CR, belonging to the SR or being uninsured. Let j represent the treatment, i.e $j \in \{CR, SR, Uninsured\}$, then let $W_{i,j} = 1$ if the individual i was exposed to treatment j. For instance, $W_{i,CR} = 1$ identifies an individual who was affiliated to the Contributive Regime, while $W_{i,CR} = 0$ implies that the individual i was not affiliated to the Contributive Regime (then, individual i was either affiliated to the Subsidized Regime or Uninsured). Z_i are the "pretreatment variables" that are not affected by the treatment (Imbens and Lemieux, 2007) like age of the head of household, age composition of the household, gender, household size and occupation status. Finally, X_i is the centered SISBEN score around zero ($c = 0$), on the threshold where the SISBEN level cross from 2 to 3 (See Figure 2).

A critical part of the Fuzzy Regression discontinuity method is the bandwidth selection. Given that the FRD method estimates the "intention to treat" in a vicinity of the threshold c, local polynomial regressions are needed to be conducted on each side of c. For this porpoise a bandwidth[25] of data should be selected to adjust said regression on each side of the cutoff c, allowing for different slopes and intercepts on either side.

Calonico et.al (2014a, 2014b) show the various options to choose said bandwidth. Lets assume that a bandwidth of size h is selected then $c - h \leq X_i \leq c + h$. The first bandwidth selection option is based on the conventional estimator for the treatment effect in the FRD. Nevertheless the FRD estimator in this case is biased and this bias is translated to the bandwidth and confidence intervals (Calonico et.al, 2014a). The second option constructs confidence intervals that are bias corrected using MSE-optimal bandwidth choices, at the cost of efficiency (Calonico et.al, 2014a), leading to wider bandwidth choices and confidence intervals. The third option is proposed by Calonico et.al (2014a, 2014b) and is called "robust". They present an alternative asymptotic approximation at the MSE bias correction, taking into account the variability of the original FRD treatment-effect estimator and the variability of the bias-correction term in the distributional approximation of the Studentized test (Calonico et.al, 2014b). Therefore this estimator is more efficient than the second one and correct the bias in the confidence intervals.

[25] A bandwidth is the interval around c, of size h, to fit a linear regression model.

For this reason, this work present two types of estimators when reporting the results: first, the conventional estimator with conventional variance estimator and bandwidth selector and, second, a "robust" estimator following Calonico et.al (2014a) which includes bias-corrected FRD estimates with MSE bias correction and the associated variance robust estimator.

5.1 Assumptions

The present section borrows from Imbens and Lemieux (2007) for giving a theoretical overview of the assumptions behind the Fuzzy Regression Discontinuity (FRD) method. Furthermore, the section explores whether the data complies with the necessary assumptions for being able to measure the effect of the health insurance type on different response variables using the FDR. Three main assumptions will be dealt with: First, the existence of a discontinuity or a jump in the treatment variable at a given value of the running variable. Second, the non-existence of a discontinuity in variables that are assumed to be exogenous. Third, the non-manipulation of the running variable (Imbens and Lemieux, 2007).

In the FRDD design, the probability of receiving the treatment changes in a proportion less than one at the threshold (Imbens and Lemieux, 2007). In other words, the assignment to the treatment group is not perfectly correlated with the crossing of the threshold:

$$lim_{x\downarrow c} Pr(W_i = 1|X_i = x) \neq lim_{x\uparrow c} Pr(W_i = 1|X_i = x) \quad (Eq.4).$$

Let the change in a regression of the outcome variable (Y) on the assignment variable (X) at each side of the threshold (c) be $lim_{x\downarrow c}\mathbb{E}[Y|X = x] - lim_{x\uparrow c}\mathbb{E}[Y|X = x]$. Analogously, the change in a regression of the treatment variable (W) on the assignment variable (X) at each side of the threshold (c) be $lim_{x\downarrow c}\mathbb{E}[W|X = x] - lim_{x\uparrow c}\mathbb{E}[W|X = x]$. Imbens and Lemieux (2007) propose to estimate the "average causal effect of the treatment" as the quotient between change in the regression of the outcome variables as numerator and the change in the regression on the assignment variable as the denominator. That is to say:

$$\tau_{FRD} = \frac{lim_{x \downarrow c} \mathbb{E}[Y|X=x] - lim_{x \uparrow c} \mathbb{E}[Y|X=x]}{lim_{x \downarrow c} \mathbb{E}[W|X=x] - lim_{x \uparrow c} \mathbb{E}[W|X=x]}$$

Imbens and Lemieux (2007) also present the estimation and inference methods associated with the Fuzzy Regression Discontinuity design. For the estimation method, they present the local linear regression method along with the bandwidth selection method associated with this local linear regression and develop the asymptotic variance estimator.

The assumptions are first stated formally in the way proposed by Imbens and Lemieux (2007), and then explored from the data perspective.

Assumption 5.1.1 (Continuity of conditional Regression Functions)

$\mathbb{E}[Y(0)|X = c] = lim_{x \uparrow c} \mathbb{E}[Y(0)|X = x] = lim_{x \uparrow c} \mathbb{E}[Y(0)|W = 0, X = x] = lim_{x \uparrow c} \mathbb{E}[Y|X = x]$

As Lee and Lemieux (2010) pointed out, this assumption is necessary as the pair $Y_i(0), Y_i(1)$ is not observable; only $\mathbb{E}[Y(0)|X = x]$ and $\mathbb{E}[Y(1)|X = x]$ over subpopulations are comparable. Such comparison is only possible if the expected values of the outcomes are continuous around the threshold. In Lee and Lemieux (2010) words "this continuity condition allows to use the average outcome of those right below the cutoff [...] as a counterfactual for those right above the cutoff". Such continuity assumption is necessary, so there are no tests for the validity of the design more than applying the FRD design itself. Lee and Lemieux (2010) also recognizes that it is not clear which behavioural assumptions on the economic agents involved are required to justify the continuity assumption around the threshold c.

Assumption 5.1.2 $W_i(x)$ is non increasing in x at $x = c$

This assumption is cited and explained by Imbens and Lemieux (2007). Basically, it is the same assumption used in the general IV setting representing monotonicity. For explaining the intuition it is important to recall that in the IV framework, the defiers are those individuals that would not take the treatment if they were assigned to the treatment group and individuals who would take the treatment if they were assigned to the control group. The monotonicity assumption in 5.1.2 implies that there are no defiers in the population (Oldenburg et. al, 2016).

Imbens and Lemieux (2007) propose to test this theoretical assumptions by means of graphical analysis. First, based on equation 4, the main identification assumption will be proved: the presence of a jump in the probability of being assigned to the treatment (W) as function of the running variable (X) on the threshold ($x = c$). Second, the continuity of the pretreatment and exogenous variables (Z) at the threshold ($x = c$) of the assignment variable (X) will be tested for identifying "possible specification problems" (Imbens and Lemieux, 2007). Third, the no manipulation of the running variable (X) will be tested examining its density and running the manipulation testing procedure proposed by Cattaneo, Jansson and Ma (2018).

Figure 3 and appendix 2 try to identify the presence of a jump of the treatment variable (belonging to the Subsidized or Contributive regime or being uninsured) as a function of the running variable (the SISBEN score) on the urban areas. The figure 3 shows the probability of having each type of insurance as a function of the SISBEN score. The top left graphic shows the result of this exercise for the subsidized regime. It shows that there is a jump down of around 20%, of the probability of being affiliated to the subsidized regime around zero. Such jump is implied by the non overlapping intervals around both sides of the threshold. Besides this, there is not only a jump in the probability of being affiliated to the subsidized regime but also a change on the trend, which changes from being decreasing, to almost flat above the zero threshold. The top left graph shows the same estimation for the contributive regime. In this case, there is also a jump upwards on the probability of being affiliated to the contributive regime of around 15%, again with non overlapping intervals. Furthermore, there is also a change in trend around the zero threshold. Finally, the left bottom graph repeats the exercise for the uninsured individuals. In this case, there is not an statistical significant jump on the probability of being uninsured around the threshold.[26]

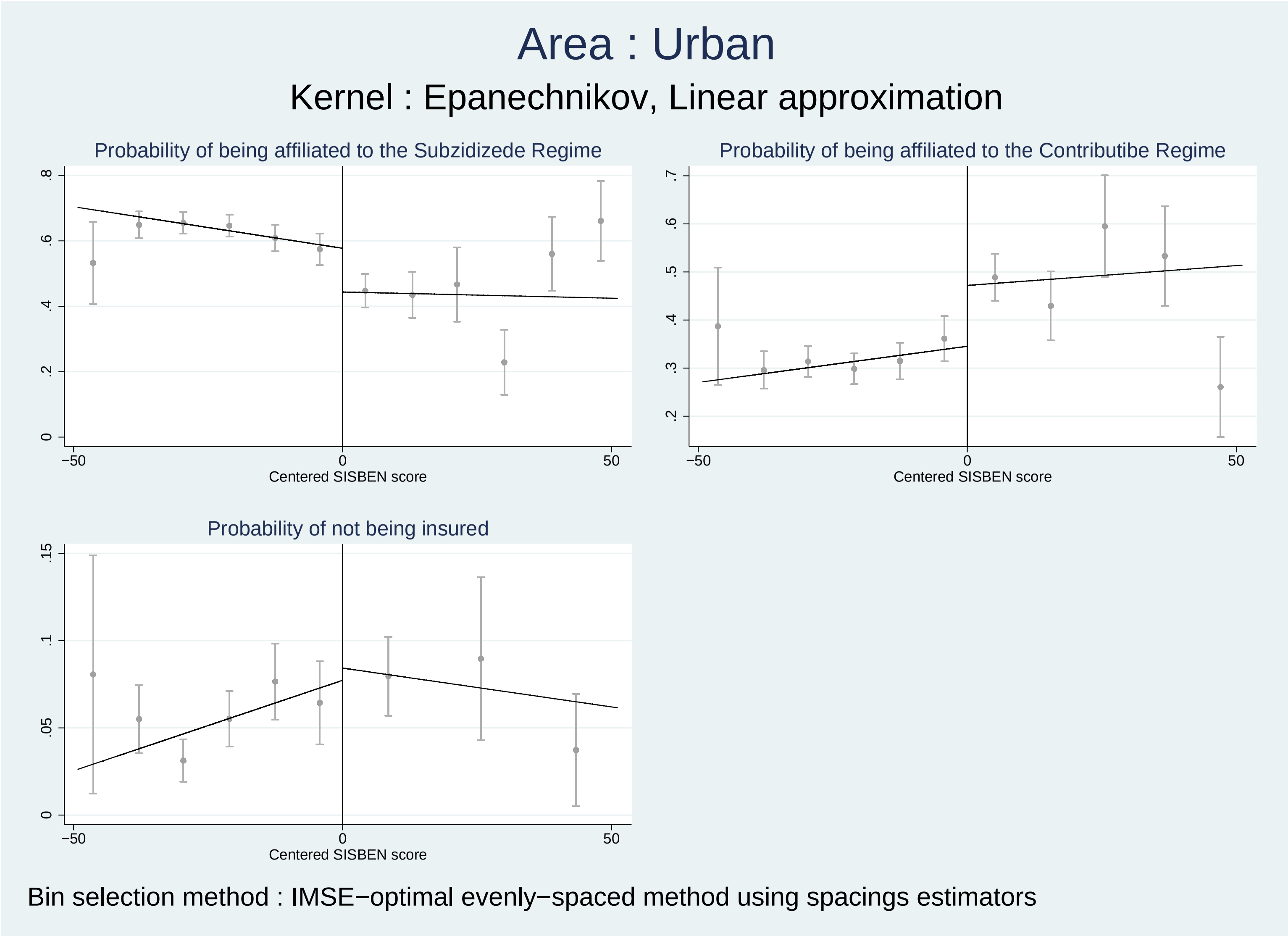
[26]The size of the jump in the affiliation to the Subsidized Health Regime is similar to the one in Miller et. al, 2013 (20%) and Camacho and Conover (2013) (15%). Additionally, various institutional arrangements make the classification of individuals among Subsidized and Contributive regime not a one to one occurrence with the SISBEN score. First, if an individual has formal employment he should be affiliated to the Contributive Regime irrespectively of the SISBEN score of the household. Second, indigenous communities and workers of public kindergartens are automatically affiliated to the Subsidized Regime irrespectively of their SISBEN score. After 2009, households belonging to the SISBEN level 3 started being eligible for the Subsidized Health Regime receiving partial services (before 2009 only SISBEN levels 1 and 2 had access to the Subsidized Regime).

The appendix 2A to 2C repeat the figure 3, but using different bandwidth selection methods, as the wideness of the intervals is very sensitive to the choice of bandwidth. Appendix 2A shows this exercise for the Subsidized Regime in urban areas, appendix 2B for the Contributive Regime in urban areas and the appendix 2C for the uninsured individuals in urban areas. In general, it can be said that the magnitude of the jump is similar in every case, despite the significance of such jump being different according to the bandwidth selection method.

Appendixes 2D to 2F repeat the exercise of appendixes 2A to 2C for rural areas. The graphs show in general that around the zero threshold there is no statistically significant jump in the probability of either being in the subsidized regime, contributive regime or uninsured. Nevertheless, in some cases there is a change in the trend around such threshold, every time on the expected direction.

Additionally, there were enrolment in the Subsidized Health Regime prior to the SISBEN survey baseline (Miller et. al, 2013). In terms of classic measurement error, Miller et al. (2013) argue that errors in household classification, manipulation of SISBEN scores and lack of public funding use to be a problem in the SISBEN I. For the other versions of the SISBEN, this may be reinforced by a lack of frequent update of the SISBEN score through a resurveying process, reporting errors in the variables on the CLSA survey and errors due to the change in the sample design of the CLSA (see Appendix 1). Finally, some gate-crashers were detected in the lowest levels of SISBEN for getting benefits from various social programs.

Figure 3: Discontinuity identification: All health insurance regimes



Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the health insurance status. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: MSE optimal

Appendixes 3A to 3C estimate the existence of a jump for the exogenous variables that should have no direct influence on the probability of being affiliated to a particular insurance regime, as function of the running variable. In general, most of the exogenous variables do not show a statistically significant jump on the SISBEN score at the zero threshold. Specifically we look at variables including age, age composition of the household, years of education, income, gender, occupation and economic status of the members of the household. On the one hand, among those variables that do not present an statistically significant jump (using a linear approximation and an epanechnikov kernel) are the age of the head of the household, the proportion of the members of the household between 0 and 12 years old, proportion of 62+ years old household members, the gender of the head of the household, the proportion of the people who do not earn income and whether the person is employed, unemployed or inactive. On the other hand, the variables that present a jump are those that are used to calculated the SISBEN score. In this case, there is a jump downwards in the the household size above the zero threshold on the SISBEN score, while the direction of the jump is upwards above the zero threshold for the years of education, per capita income and the proportion of wage earners among household members.

Table 2 repeats the same exercise but estimating the descriptive statistics of the exogenous variables for each health insurance type. In this particular case, variables related to age (age of the head of the household, proportion of members of the household between 0-12 years old and proportion of members of the household older than 62 years old) seem to be balanced according to the treatment status. In other words, those variables are similar irrespective of whether the household or the individual belongs to the Contributive or Subsidized Regimes, or it is uninsured. The same can be said with respect to gender of the head of the household and household size.

Other socio economic status variables, seem to be unbalanced according to the insurance status. This may be due to the fact that some of those variables are used for either calculating the score of the running variable (SISBEN score) or because they determine directly the insurance status in some cases (like being an employee or unemployed). Here, the members of the households belonging to the Contributive Regime seem to have around three more years of education (12.08 years) when compared with the members of the households belonging to the Subsidized Regime (8.93 years). The members belonging to the uninsured households are in the middle between the last two (9.8 years of education). The same pattern is present in the average per-capita income divided by the poverty line as the households belonging to the Subsidized Regime have a much lower income on average when compared with households in the Contributive Regime (0.81 for the SR versus 1.41 for the SR). In the same direction, the household per capita income of the uninsured households are in the middle between the last two (0.91). In respect to the occupation status, 67% of members of the households belonging to the CR are wage earners, while 79% of the members of the households belonging to the SR are self-employees. Finally, 67% of the members belonging to households in the CR are employed, while this percentage decreases to 55% among members of households belonging to the SR. Similarly, the unemployment rate tends to be higher among members of households belonging to the SR.

Table 2: Descriptive statistics for the exogenous variables of all individuals included in econometric models, Urban areas

PANEL E : EXOGENOUS VARIABLES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SUBSIDIZED REGIME			CONTRIBUTIVE REGIME			UNINSURED		
VARIABLES	mean	sd	N	mean	sd	N	mean	sd	N
Age [2010, 2013 and 2016]	28.43	19.52	2,360	30.72	18.46	1,211	30.09	19.01	194
Proportion of members between 0 and 12 years old [2010, 2013 and 2016]	0.308	0.210	2,360	0.272	0.201	1,211	0.262	0.210	194
Proportion of members 62+ years old [2010, 2013 and 2016]	0.0588	0.131	2,360	0.0575	0.139	1,211	0.0582	0.130	194
Education [2010, 2013 and 2016]	8.937	4.646	1,440	12.08	4.234	815	9.878	4.736	131
Per-capita income divided by poverty line [2010, 2013 and 2016]	0.814	0.630	2,360	1.417	1.063	1,211	0.915	0.899	194
Gender: Male [2010, 2013 and 2016]	0.458	0.498	3,182	0.472	0.499	1,620	0.505	0.501	289
Household size [2010, 2013 and 2016]	5.284	2.129	2,360	4.789	1.817	1,211	5.093	2.326	194
Occupation status: Wage earner [2010, 2013 and 2016]	0.167	0.373	1,303	0.674	0.469	870	0.222	0.417	135
Occupation status: Self-employed [2010, 2013 and 2016]	0.790	0.408	1,303	0.317	0.466	870	0.719	0.451	135
Occupation status: No income earner [2010, 2013 and 2016]	0.0437	0.205	1,303	0.00920	0.0955	870	0.0593	0.237	135
Economic Status: Employee [2010, 2013 and 2016]	0.552	0.497	2,361	0.672	0.470	1,294	0.659	0.475	205
Economic Status: Inactive [2010, 2013 and 2016]	0.00847	0.0917	2,361	0.00464	0.0680	1,294	0.00488	0.0698	205
Economic Status: Unemployed [2010, 2013 and 2016]	0.440	0.496	2,361	0.323	0.468	1,294	0.337	0.474	205

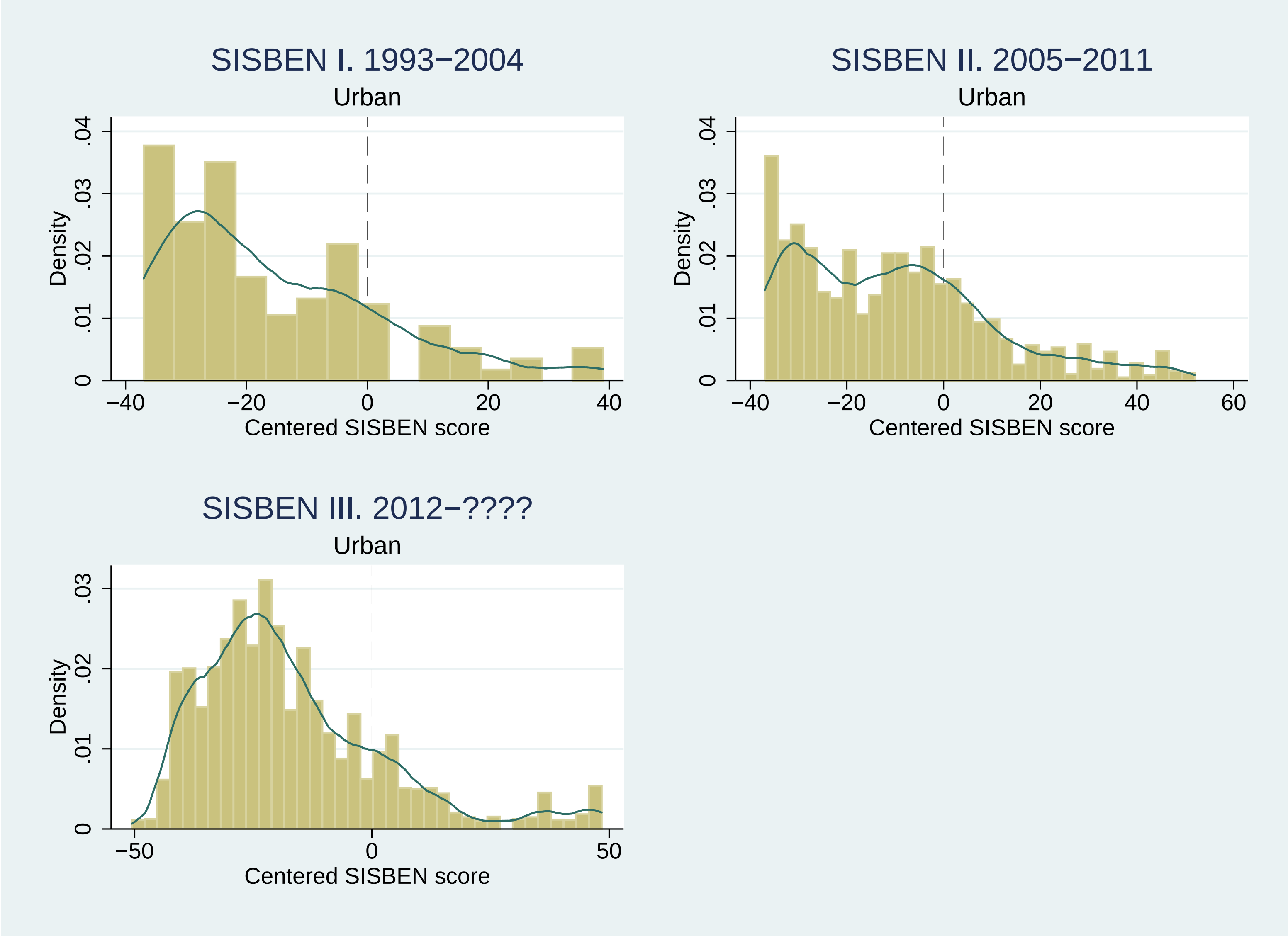
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Sample divided according to Health Insurance Status 2./ Restricting the sample only to individuals in the econometric models 3./ Urban households only

Finally, figure 4 and appendix 4A and 4B try to identify the existence of manipulation in the running variable (the SISBEN score) for the three phases of the SISBEN. Camacho and Conover (2011) have documented well the existence of manipulation of the SISBEN score phase I, as both, households and local authorities had the incentives to include more households on the subsidized health regime plus the fact that the algorithm to calculate the SISBEN phase I score was released to the public, allowing them to manipulate certain variables in order to be included in the subsidized health regime. What the authors found is a strong accumulation of observations just before the cuttoff point where the households are eligible for the subsidized regime, and a huge fall in the number of observations just after such point. They use administrative data of the SISBEN score. Figure 4 shows the histogram and kernel density estimates for the SISBEN scores on its three different phases for the urban areas. The top left graph on figure 4 shows the density for the SISBEN phase I. Here, the density of the score shows a certain anomaly before the centred cutting point of zero. The kernel density estimates have a flat zone, while the histogram presents a spike just before zero. The top left graph shows the density of the SISBEN score phase II. Here the kernel estimate is much flatter respect to the SISBEN I score, and there is a peak around the zero threshold. Here, it is not clear whether there is evidence of manipulation on this version of the SISBEN, as the peak is before and after zero. The same is true for the SISBEN III, where the evidence of such an anomaly is even lower.

Appendix 4A repeat the same exercise for rural areas, and it appears to show less evidence for the manipulation of the SISBEN score, even for the SISBEN score phase I. In any case, only 4.7% of the individuals in the database report having the SISBEN score phase I as the current SISBEN score in 2013, while 58.8% of the individuals are part of the SISBEN score phase II and 36% report having as valid score, the SISBEN phase III. Appendix 4B show the running variable manipulation test developed by Cattaneo et.al (2017) and implemented by Cattaneo et.al (2018). The density graphs of the running variable indicate that there is some evidence of SISBEN score manipulation for the SISBEN I and III for the urban areas. To confirm this, the corresponding statistical test was conducted. Under the null hypothesis of said test, there is no manipulation in the running variable. The corresponding p-values in the urban areas are 0.0027 for the SISBEN I, 0.32 for the SISBEN II and 0.88 for the SISBEN III.

This implies that in case of having any manipulation of the SISBEN score, it would be driven by the SISBEN phase I, which represent only a small fraction of the sample and therefore can be ignored safely. As a robustness check for our econometric results, the models were also run ignoring individuals who are part of the SISBEN phase I score, and the results remain unchanged.

Figure 4: Urban SISBEN scores in each phase. Colombian Longitudinal Survey by Universidad de los Andes



Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Histogram and Kernel density estimates of the SISBEN score by household. 2./ SISBEN I, II and III separately and centered around zero. 3./ Restricting the sample only to households with individuals in the econometric models. 4./ Epanechnikov kernel

5.2 Empirical predictions of the effect of different types of health insurance

Health condition variables are the output on the health production function, while the preventive and curative services work as inputs in that production function. Such inputs in turn depend on the type of insurance the individual has in terms of quality and reach, so they should be analysed through their particular effect among CR, SR and uninsured status. Medical care use variables can be affected throught preventive services and curative services. For the preventive services, they are free in the CR, SR and even for the uninsured individuals. At first sight there should be no difference in the use of preventive healthcare services from the demand side, while the individuals on the SR may have less incentives to use the preventive services, as the curative services are free for them. For the curative health services the SR has a reduced price compared with the CR and the uninsured individuals which should increase their use on the SR. In any case, the uninsured individuals pay the more expensive curative healthcare services, so I expect a lower use of curative services on this population.

One of the main purposes of the health insurance is to protect households against the medical care cost associated with unexpected illness. In that sense, the individuals on the SR and CR are expected to have lower out of pocket medical expenses per episode of illness in comparison with those expenses for the uninsured individuals. Pointing towards the same direction, all of the other household expenses of the insured households should suffer less when confronting an unexpected illness. The behavioural distortions may be triggered indirectly by the prices faced by the individuals according to the health insurance type they have. For instance, uninsured individuals may have more incentives for following a healthy style of life as they face higher prices for the curative healthcare services, while individuals on the SR would have less incentives for having such healthy life style.

5.3 Results

This section presents the results of the econometric models for evaluating the effect of the different types of health insurance on various response variables. This is done using the Fuzzy Regression Discontinuity method combined with instrumental variables in a 2SLS setting as explained in section 5.

In the first stage, a regression of the treatment variable (either belonging to the Subsidized Regime, Contributive Regime or being uninsured) is run using as exogenous variable the SISBEN score. Then, in the second stage, and assuming that the assignation on any of the treatments is random in a vicinity around the zero threshold of the centred SISBEN score, a regression of the response variable (for instance health status variable) on the prediction of the health insurance status variable in the first stage is run. Previous works have used as treatment group the individuals belonging to the Subsidized Health Regime and as control groups any other health insurance alternative (belonging to the Contributive Regime or being uninsured) (see Camacho and Conover, 2013 and Miller, et. al, 2013). Here I run the models for the three possible treatments using as controls the other two insurance types.

In other words, in this paper, when evaluating the Subsidized Regime I use as control group a combination of the Contributive Regime and uninsured individuals. When evaluating the Contributive Regime, I use as control group a combination of the individuals belonging to the Subsidized Regime and those who are uninsured. Finally, when evaluating the uninsured group as treatment the control group is composed by a combination of the Contributive and Subsidized Regimes. Tables 3 to 6 present the results of such econometric models

Columns 1 and 2 of tables 3 to 6 show the results of the fuzzy regression discontinuity model for the Subsidized Regime as treatment. Column 1 present the conventional Fuzzy Regression Discontinuity estimators with the conventional standard error, while column 2 presents the bias-corrected Fuzzy Regression Discontinuity estimates with standard error estimators that are asymptotically robust as explained in section 5.1. Columns 3 and 4 repeat the same procedure, but using as treatment the Contributive Regime, while columns 5 and 6 do the same using as treatment group the uninsured individuals. Column 7 shows the number of observations.

It is important to note that the SISBEN score was only asked as a retrospective question in 2013 when people was inquired about the score and the year when the last SISBEN interview was conducted to the household. With this information I can assume with precision which version of the SISBEN score the household has (I, II or III) for the years before 2013. For the year 2016 I assume that the SISBEN score is the same than in 2013. This assumption, despite being strong, do not alter the results importantly.

In general, the results show that there are no significant differences in any of the response variables evaluated here between the Subsidized Health Regime, the Contributive Health Regime or being uninsured. This result contradicts the findings of Camacho and Conover (2013) and Miller et. al (2013) who found that up to 2005 still existed statistical significant differences between being affiliated to the Subsidized Regime and a control group composed by a combination of households affiliated to the Contributive Regime and uninsured households. Nevertheless, the results on this paper go in the direction of one of the goals of the Colombian Ministry of Health, which was closing the gap in quality, use, and health outputs among the households affiliated to the Subsidized Regime and those affiliated to the Contributive Regime. Furthermore, the results in this work confirm the results found by Econometria (2015) who found equalization of the Contributive and Subsidized Health insurance regimes in a wide variety of indicators using differences in differences models for the years 2011 and 2013.

5.3.1 Health status

Table 3 presents the results of being affiliated to different insurance regimes on various health outputs. With respect to the circulatory system diseases, the affiliated to the Contributive Health Regime seems to have less incidence of circulatory system diseases as most of the coefficients for those type of conditions are negative. Meanwhile, the individuals who are either, affiliated to the Subsidized Regime or uninsured show a higher incidence of circulatory system diseases. Among the Subsidized Regime affiliates and the uninsured individuals, the former show a higher incidence of circulatory system diseases. In any case, none of the coefficients are statistically different from zero, indicating that from an econometric perspective there is no significant difference in the incidence of circulatory system diseases among any of the treatment and control groups.

When considering the incidence of chronic diseases, those individuals affiliated to the Subsidized and Contributive Health Regimes seem to have a lower incidence of chronic diseases, while those uninsured have a higher incidence compared with their respective control group. Again, none of the coefficients are statistically different from zero, implying that there is no statistical difference among the treatment and its respective control groups.

The only information the Colombian Longitudinal Survey by Universidad de los Andes (CLSA) has on infectious diseases is tuberculosis. In this case, the coefficients for the three insurance groups are very small (practically zero) indicating no difference in the incidence of such condition between treatment and control groups.

The uninsured individuals show a higher incidence of conditions that affect the activities of daily living (ADL's), while those individuals affiliated to the Contributive Regime show the lowest incidence among the three control groups. Again, none of the estimated coefficients are statistical different from zero on any combination of treatment and control groups.

Finally, the results concerning health events during last year are more mixed. Individuals on the Subsidized Health Regime seem to have a higher incidence of illness, surgery and pregnancy, while individuals on the Contributive Regime have more incidence of accidents, and dentistry events according to the robust results. Uninsured individuals have a higher incidence of surgery. On the other side of the spectrum individuals affiliated to the Subsidized Regime have the lowest incidence of unexpected accidents, while the uninsured individuals have the lowest incidence of unexpected pregnancy and dentistry events. Again, none of the coefficients are statistically different from zero.

Table 3. Fuzzy Regression Discontinuity results: Health Status

	Subsidized Regime		Contributive Regime		Uninsured		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Conventional	Robust	Conventional	Robust	Conventional	Robust	Observations
CIRCULATORY SYSTEM DISEASE - 2013 and 2016							
Thrombosis	-0.319	-0.131	0.149	0.108	0.436	0.374	1783
	(0.428)	(0.489)	(0.168)	(0.189)	(0.728)	(0.809)	1783
Hearth attack	0.143	0.0826	-0.110	-0.101	-0.269	-0.350	1783
	(0.089)	(0.102)	(0.326)	(0.374)	(0.724)	(0.841)	1783
Hearth condition	0.341	0.244	-0.429	-0.310	-1.535	-1.837	1783
	(0.302)	(0.361)	(0.403)	(0.466)	(2.638)	(3.156)	1783
Hypertension (Yes)	2.089	-0.00160	-1.089	-0.704	-4.720	-4.139	1783
	(2.665)	(3.251)	(1.067)	(1.209)	(7.069)	(8.300)	1783
Hypertension (Only during pregnancy)	0.479	1.356	-0.525	-1.436	0.692	0.600	1783
	(0.892)	(1.106)	(1.004)	(1.262)	(1.133)	(1.401)	1783
CHRONIC DISEASE - 2013 and 2016							
Asthma	-0.662	-0.193	2.104	-1.563	2.206	1.974	1783
	(0.627)	(0.769)	(5.371)	(6.496)	(3.478)	(4.126)	1783
Emphysema	-0.0450	0.0176	0.0585	-0.00730	0.366	0.167	1783
	(0.121)	(0.142)	(0.135)	(0.155)	(0.779)	(0.911)	1783
Diabetes (Yes)	0.953	-0.0828	-0.458	-0.389	-2.165	-2.236	1783
	(1.395)	(1.726)	(0.686)	(0.778)	(3.550)	(4.161)	1783
Diabetes (No)	-0.481	-0.426	0.644	0.582	2.925	3.280	1783
	(0.532)	(0.600)	(0.764)	(0.858)	(4.074)	(4.795)	1783
Diabetes (Only during pregnancy)	-0.935	-3.940	-0.344	-0.235	-0.831	-0.968	1783
	(3.380)	(4.045)	(0.569)	(0.668)	(1.446)	(1.760)	1783
Ulcer	-0.135	-0.329	0.0582	0.0659	0.152	0.575	1783
	(1.488)	(1.710)	(0.453)	(0.536)	(1.494)	(1.730)	1783
Epilepsy	0	0.00170	0	-0.00180	-0.241	-0.266	1783
	(0.000)	(0.003)	(0.000)	(0.004)	(0.358)	(0.387)	1783
Cancer	0.0488	0.0354	-0.0507	-0.0374	0.0978	0.0757	1783
	(0.059)	(0.068)	(0.071)	(0.082)	(0.378)	(0.479)	1783
INFECTIOUS DISEASES - 2013 and 2016							
Tuberculosis	-0.00980	-0.0103	0	-0.00320	0	0.0498	1783
	(0.011)	(0.015)	(0.000)	(0.004)	(0.000)	(0.036)	1783
ACTIVITIES OF DAILY LIVING (ADL) - 2010, 2013 and 2016							
Movement difficulties	0.0145	0.0168	-0.0130	-0.0144	-0.550	-0.883	4149
	(0.062)	(0.074)	(0.056)	(0.067)	(2.156)	(2.415)	4149
Showering difficulties	-0.00380	-0.00170	-0.00890	-0.00950	0.172	0.189	4149
	(0.022)	(0.024)	(0.017)	(0.020)	(0.452)	(0.493)	4149
Learning disability	0.0204	0.0368	-0.0516	-0.0577	0.133	0.170	4238
	(0.039)	(0.042)	(0.036)	(0.038)	(0.232)	(0.271)	4238
Blindness	0.0805	-0.122	-1.273	54.2819**	-0.765	-0.950	4238
	(0.101)	(0.115)	(13.963)	(21.414)	(3.272)	(4.221)	4238
HEALTH EVENT - 2010, 2013 and 2016							
Illness	-8.198	45.29	-0.0469	-0.0941	-0.0841	0.109	4084
	(58.122)	(68.746)	(0.396)	(0.445)	(1.165)	(1.317)	4084
Accident	-0.127	-0.0779	0.0768	0.0714	-0.0727	-0.0447	4084
	(0.164)	(0.191)	(0.091)	(0.106)	(0.236)	(0.278)	4084
Dentistry	-1.138	0.0816	0.423	0.261	-1.194	-0.998	4084
	(1.541)	(1.802)	(0.282)	(0.339)	(0.943)	(1.050)	4084
Surgery	0.189	0.0290	-0.0861	-0.0728	0.221	0.171	4084
	(0.290)	(0.360)	(0.098)	(0.118)	(0.231)	(0.270)	4084
Pregnancy	-0.0824	0.00350	0.0180	-0.00570	-1.003	-0.924	1449
	(0.095)	(0.097)	(0.020)	(0.013)	(3.072)	(3.436)	1449

Source: Author's estimations using CLSA for the years 2013 and 2016 or 2010, 2013 and 2016 when indicated.
Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

5.3.2 Behavioral distortions

Table 4 present the fuzzy regression discontinuity results for the behavioural distortion variables. Individuals affiliated to the Subsidized Health Regime tend to have a higher frequency of fruit consumption per week, while uninsured people tend to have a lower fruit consumption per week when comparing the robust coefficients. None of the estimated coefficients are statistically different from zero.

Individuals affiliated to the Contributive Regime tend to have a higher frequency of vegetable consumption, while those in the Subsidized Regime tend to be concentrated in either a high frequency consumption or a very low consumption of vegetables per week. The uninsured individuals tend to consume less vegetables compared with the insured individuals. Although those results, there is not a clear pattern in the vegetable consumption among insurance regimes and there are no statistical differences among any of the treatment and control groups.

With respect to the consumption of fried food, individuals affiliated to the Contributive Regime and uninsured individuals tend to have a higher frequency of consumption of fried food per week, while individuals on the Subsidized Regime tend to be concentrated in the consumption of fried food in a low frequency. There is no statistically significant difference among groups.

Finally, the physical activity is evaluated as moderate and strong physical activity in the last seven days, the number of days during the week and the number of minutes per day. The uninsured individuals tend to do physical activity less frequently, compared with its control groups, and also spend less time doing such physical activity. On the other end, individuals affiliated to the Subsidized Regime tend to do more strong physical exercise during the last week for a longer duration, while people on the Contributive Regime tend to do moderate exercise more days on the week, with respect to its controls. None of these results are statistically significant.

Table 4. Fuzzy Regression Discontinuity results: Behavioral distortions

	Subsidized Regime		Contributive Regime		Uninsured		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Conventional	Robust	Conventional	Robust	Conventional	Robust	Observations
FRUIT CONSUMPTION - 2013 and 2016							
Less than once a week	-0.0987	0.0643	0.196	0.0237	0.986	0.465	1783
	(0.307)	(0.369)	(0.376)	(0.430)	(1.430)	(1.656)	1783
Once a week	0.252	0.115	-0.339	-0.124	-0.746	-0.595	1783
	(0.412)	(0.500)	(0.589)	(0.704)	(1.889)	(2.254)	1783
2-4 times a week	-0.0783	0.153	0.00660	-0.110	0.976	-0.300	1783
	(0.494)	(0.592)	(0.582)	(0.650)	(2.841)	(3.437)	1783
5-6 times a week	0.366	0.128	-0.558	-0.106	-1.766	-1.339	1783
	(0.484)	(0.575)	(0.801)	(0.954)	(2.899)	(3.489)	1783
Daily	-0.705	-0.702	1.146	0.906	3.351	5.160	1783
	(0.607)	(0.653)	(1.389)	(1.667)	(5.056)	(6.111)	1783
More than once a day	1.386	0.303	-0.672	-0.648	-2.643	-3.118	1783
	(2.314)	(2.847)	(0.842)	(0.929)	(4.287)	(4.951)	1783
VEGETABLE CONSUMPTION - 2013 and 2016							
Less than once a week	0.256	0.252	-0.345	-0.302	-1.499	-2.078	1783
	(0.232)	(0.265)	(0.349)	(0.393)	(2.261)	(2.669)	1783
2-4 times a week	-0.321	0.488	-1.203	-2.062	0.607	-0.782	1783
	(1.104)	(1.363)	(2.039)	(2.386)	(3.210)	(3.872)	1783
5-6 times a week	-0.414	-0.322	4.038	-11.58	2.058	2.127	1783
	(0.534)	(0.624)	(19.747)	(24.318)	(3.538)	(4.291)	1783
Daily	-3.865	-12.18	-0.0240	0.124	-3.244	-3.062	1783
	(8.937)	(10.754)	(0.950)	(1.094)	(5.463)	(6.378)	1783
More than once a day	0.777	0.186	-2.795	2.077	-2.326	-1.736	1783
	(0.887)	(1.086)	(8.966)	(10.418)	(3.784)	(4.547)	1783
FRIED FOOD CONSUMPTION - 2013							
Less than once a month	0.318	0.169	-0.365	-0.00660	-1.322	-1.160	1547
	(0.288)	(0.332)	(0.395)	(0.467)	(2.273)	(2.662)	1547
Once a month	-0.432	-0.392	0.565	0.408	2.035	2.657	1547
	(0.446)	(0.504)	(0.749)	(0.877)	(3.339)	(3.805)	1547
2-3 times a month	-0.652	-0.632	0.629	0.591	2.697	3.187	1547
	(0.535)	(0.604)	(0.513)	(0.556)	(4.158)	(4.604)	1547
Twice a week	1.839	1.375	-4.676	-1.534	-7.987	-11.27	1547
	(1.487)	(1.727)	(13.431)	(14.892)	(12.858)	(15.098)	1547
3-4 times a week	0.128	0.0543	-0.149	-0.0421	-2.281	-2.013	1547
	(0.718)	(0.827)	(0.901)	(1.039)	(4.260)	(5.117)	1547
5-6 times a week	0.199	0.185	-0.317	-0.293	-0.690	-0.795	1547
	(0.340)	(0.390)	(0.612)	(0.692)	(1.666)	(1.956)	1547
Daily	-1.247	-0.907	2.153	0.500	5.498	5.896	1547
	(0.974)	(1.107)	(2.405)	(2.889)	(9.270)	(10.962)	1547
Twice a day	-0.351	-0.194	0.452	0.170	0.665	0.814	1547
	(0.394)	(0.479)	(0.592)	(0.718)	(1.165)	(1.381)	1547
Three or more a day	0.0740	0.172	0.104	-0.0533	0.178	0.162	1547
	(0.092)	(0.114)	(0.348)	(0.393)	(0.298)	(0.334)	1547
PHYSICAL ACTIVITY - 2013 and 2016							
Moderate: Yes in the last 7 days	-0.106	-0.121	0.0940	0.120	-0.696	-0.495	1783
	(0.247)	(0.287)	(0.360)	(0.435)	(1.741)	(2.020)	1783
Moderate: Number of days in the last 7 days	0.466	0.626	5.439	36.32	-1.838	-2.218	1783
	(2.844)	(3.488)	(43.924)	(52.910)	(6.510)	(7.975)	1783
Moderate: Minutes per day	14.38	5.247	-29.03	-2.959	-89.09	-78.47	1778
	(43.039)	(50.635)	(76.234)	(90.948)	(192.353)	(222.012)	1778
Strong: Yes in the last 7 days	0.210	0.165	-0.436	-0.299	-1.963	-1.961	1783
	(0.157)	(0.180)	(0.376)	(0.436)	(2.876)	(3.394)	1783
Strong: Number of days in the last 7 days	0.166	0.175	-0.279	-0.265	-1.144	-1.425	1783
	(0.129)	(0.156)	(0.329)	(0.416)	(1.920)	(2.311)	1783
Strong: Minutes per day	49.82	20.69	-87.72	-33.05	-342.0	-357.2	1777
	(44.883)	(54.581)	(84.765)	(103.228)	(531.586)	(646.720)	1777

Source: Author's estimations using CLSA for the years 2013 or 2013 and 2016 when indicated.
Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

5.3.3 Medical Care Use

As with the other variables, there is no clear pattern for which individuals use more the preventive medicine services according to the insurance type. In particular, individuals affiliated to the Subsidized Regime tend to have less hospitalizations in the last 12 months and tend to use less contraceptives, less use of dentistry and optometrist services. Individuals affiliated to the Contributive Regime tend to go more to the dentist and been hospitalized in the last 12 months. The uninsured individuals tend to visit the doctor more, but use all the other services less. The subsidized regime affiliates use less contraceptives. Again, none of the results are statistically different from zero implying that there is no differences among the treatment and control groups.

Table 5. Fuzzy Regression Discontinuity results: Medical Care Use

	Subsidized Regime		Contributive Regime		Uninsured		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Conventional	Robust	Conventional	Robust	Conventional	Robust	Observations
MEDICAL CARE USE - 2010, 2013 and 2016							
Hospitalized in the last 12 months	-0.762	-0.463	0.500	0.265	-1.054	-1.025	4084
	(0.810)	(0.920)	(0.451)	(0.544)	(0.966)	(1.058)	4084
Doctor	0.490	0.263	-0.166	-0.0770	0.829	0.514	4084
	(0.819)	(0.905)	(0.434)	(0.478)	(1.431)	(1.594)	4084
Dentist	-0.971	-0.888	0.708	0.625	-1.929	-1.816	4084
	(0.926)	(1.006)	(0.600)	(0.684)	(1.720)	(1.884)	4084
Optometrist	-0.995	-1.822	-0.331	-0.0221	-0.129	-0.317	4084
	(1.998)	(2.374)	(0.809)	(0.938)	(1.105)	(1.219)	4084
Alternative Medicine	0.810	2.242	0.121	0.0834	-0.320	-0.297	4084
	(1.944)	(2.310)	(0.105)	(0.124)	(0.281)	(0.319)	4084
Contraceptive	-0.828	0.769	0.270	0.0672	-0.491	-0.428	2636
	(2.217)	(2.634)	(0.506)	(0.597)	(0.979)	(1.118)	2636
Other	-0.0806	-0.0558	0.0987	0.0591	0.572	0.371	2968
	(0.239)	(0.279)	(0.283)	(0.329)	(2.091)	(2.404)	2968

Source: Author's estimations using CLSA for the years 2010, 2013 and 2016 when indicated.

Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

5.3.4 Risk Protection and Consumption Smoothing

Table 6 show the results for the variables in the category risk protection and consumption smoothing. This includes the household monthly expenditure, household food expenditure and household annual expenditure. There is no a statistically significant effect on any of the response variables studied here due to any of the health insurance regimes. Nevertheless, in this case, such effect has the expected sign. If the Subsidized Health Regime works as a way for protecting households from financial risk generated from health shocks, it should increase the monthly expenditures in non-health items. In respect to the food expenditure the estimated coefficient has the contrary sign than expected: households belonging to the Subsidized Health Regime present a decrease in their food expenditure. Finally, for the out of pocket health annual expenditure, the estimation presents a negative sign for households in the Subsidized Health Regime making evident that the Subsidized Regime is fulfilling the task of insuring the poorest people against the financial risk of health shocks.

None of the other variables have any particular effect according to the insurance type.

Table 6: Fuzzy Regression Discontinuity results: Risk Protection and Consumption Smoothing

	Subsidized Regime		Contributive Regime		Uninsured		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Conventional	Robust	Conventional	Robust	Conventional	Robust	Observations
RISK PROTECTION AND CONSUMPTION SMOOTHING - 2010, 2013 and 2016							
Log Household total monthly expenditure	0.150	0.0473	-0.174	-0.111	2.898	1.582	5089
	(0.352)	(0.387)	(0.225)	(0.244)	(3.171)	(3.631)	5089
Log Household food monthly expenditure	-0.0484	-0.0955	-0.00310	-0.0115	-0.212	-0.830	5089
	(0.410)	(0.443)	(0.299)	(0.327)	(1.980)	(2.270)	5089
Log Household health annual expenditure	-4.117	-2.385	4.105	2.497	-68.04	-54.01	629
	(2.589)	(3.200)	(3.265)	(3.898)	(208.378)	(267.621)	629

Source: Author's estimations using CLSA for the years 2010, 2013 and 2016 or 2013 and 2016 when indicated.

Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

In general, this results support the hypothesis of the equalization of the Subsidized and Contributive Health Regime, after the government started to equalize the Mandatory Health Plans in 2009. This can be said for all the set of variables studied here: health status, behavioural distortions, medical care use and risk protection and consumption smoothing. This result implies that after more than 20 years of the implementation of the Subsidized Health Regime the users of such insurance type (the more vulnerable people) enjoy the same health status, medical care use and risk protection against illness compare to those individuals in the Contributive Regime. At a first glance such result may seem promising as a support for the claim that both, the Subsidized and Contributive Health Regime have reached an equally good status in terms of insurance against health shocks for their respective users. Nevertheless recent scandals among insurance companies belonging to the Contributive Health Regime (Saludcoop, Medimas) concerning the mismanage of the funds collected by them, jointly with the lack of payment from the insurance companies to the healthcare providers, and an increasing number of complains from the users regarding the quality of the healthcare service, decreases the likelihood of such hypothesis being true. If those cases of mismanage of the funds on the private insurance companies are true, it would not be clear if the result on this paper is due to an improvement in the quality and coverage of the Subsidized Health Regime, a decrease on the quality of the Contributive Health Regime or a combination of both. It is important also to note that there is no significant difference between being insured by any of the two health insurance regimes and being uninsured. This may be due to the fact that basic medical attention and emergency room services are warranted to the population irrespectively of their health insurance status.

Additionally, it should be pointed out the relatively small sample I am dealing with in this paper, which leads to very imprecise estimations and may result on the increase of the type II error, specially for the uninsured individuals.

Another possible explanation for the lack of significance of the results in this paper is the size shrinkage of the uninsured group. As previously mentioned Camacho and Conover (2013) and Miller et. al (2013) found a significant difference in various response variables between being insured through the Subsidized Health Regime and a control group that mixes Contributive Regime Affiliates and uninsured population. Indeed the uninsured population in Miller et. al (2013) represents 38% of the sample and 25% in Camacho and Conover (2013) while in this study it only represents 5%. This characteristic, jointly with the fact that those studies use data for the period 1998-2005, before reaching universal coverage of the healthcare system, when uninsured population still had no entitlement to any healthcare service in practice, make this hypothesis plausible. If this is the case, most of the significant effect of the Subsidized Health Regime in those papers would come from the change in the uninsured status at the threshold when uninsured individuals become eligible to the Subsidized Health Regime. Moreover, when testing the existence of a jump in the affiliation to the Subsidized Health Regime status at the designated SISBEN score threshold, Camacho and Conover (2013) found that said jump exists, but it does not exist when testing its existence on the Contributive Health Regime affiliation status.

To further explore the econometric results and its lack of significance, I graph the estimated coefficients and its respective 95% confidence interval in the figures 5 to 8. The purpose is to understand whether the lack of significant results is mainly due to imprecise estimates or the effect of the scale of the variables. On the one hand, if the lack of significance is due to high variability on the data, the confidence intervals on the coefficients will be wide irrespectively of the scale of the coefficient and will include the zero in every case, usually due to the amplitude of the interval. This imprecise estimates may be due to a small sample for some of the categories of the variables on some of the insurance regimes. On the other hand, if the confidence intervals are small, but the coefficients are also small and the interval includes the zero it may be interpreted as the treatment variable (SR, CR and uninsured status) having some effect on the response variables (Y_i) that is small, depending on how close to the center of the interval the zero is. In this case the graphs were done using standardized data (susbstracting the mean of each variable and dividing by the corresponding standard deviation) for the SISBEN score, treatment variable and response variable. Following this procedure the size of the estimated coefficients is comparable and it is possible to draw conclusions on which variables do the health insurance regime have more effect.

When using standardized data, the coefficient obtained is called standardized coefficient. It allows to compare the scale of the coefficient among regressor (in the multiple regression case), or models in this case, as long as all the involved variables are standardized. The higher the value of the coefficient estimated, the “stronger” the effect on the response variable. The reason why the effect on the response variable is comparable among standardized variables is that this process eliminates the scale of the variables allowing to interpret them as changes in standard deviations of the variables. For instance, in this case the uninsured status has a coefficient of 1.2 on the incidence of thrombosis. This means that for each standard deviation of increase in the uninsured status the incidence of thrombosis increases in 1.2 standard deviations.

The figure 5 shows the graph of the standardized coefficients for the health status variables, Y_i, and the three possible treatments $W_{i,j} = 1$ with $j \in \{CR, SR, Uninsured\}$. In this case all of the variables are precisely measured when dealing with the uninsured individuals, except the absence of diabetes and blindness. For the other variables these intervals indicate that the treatment may still have some effect on the response variables, but that it may just very small. For the Contributive Regime, the results for the diabetes status and hypertension may be attributed to the imprecise estimates as can be deducted from the wide interval. Same conclusion can be drawn for the variables hypertension, dentistry and surgery in the Subsidized Regime. The figure 6 repeats the exercise for the behavioural distortion variables. In this case most of the variables are precisely measured and include the zero in the confidence interval, with the exception of the variables related to fried food consumption for the uninsured which have wide intervals.

Figure 7 shows the same graph for medical care use variables. Most of the variables have a medium level of confidence intervals for the three regimes, while only the variables hospitalized in the last 12 months and visit to the optometrist for households in the Subsidized Regime and other for the uninsured individuals present a wide confidence interval. Figure 8 repeats the exercise for the risk protection and smoothing consumption variables. In this case only the household health annual expenditure shows wide intervals for the uninsured individuals.

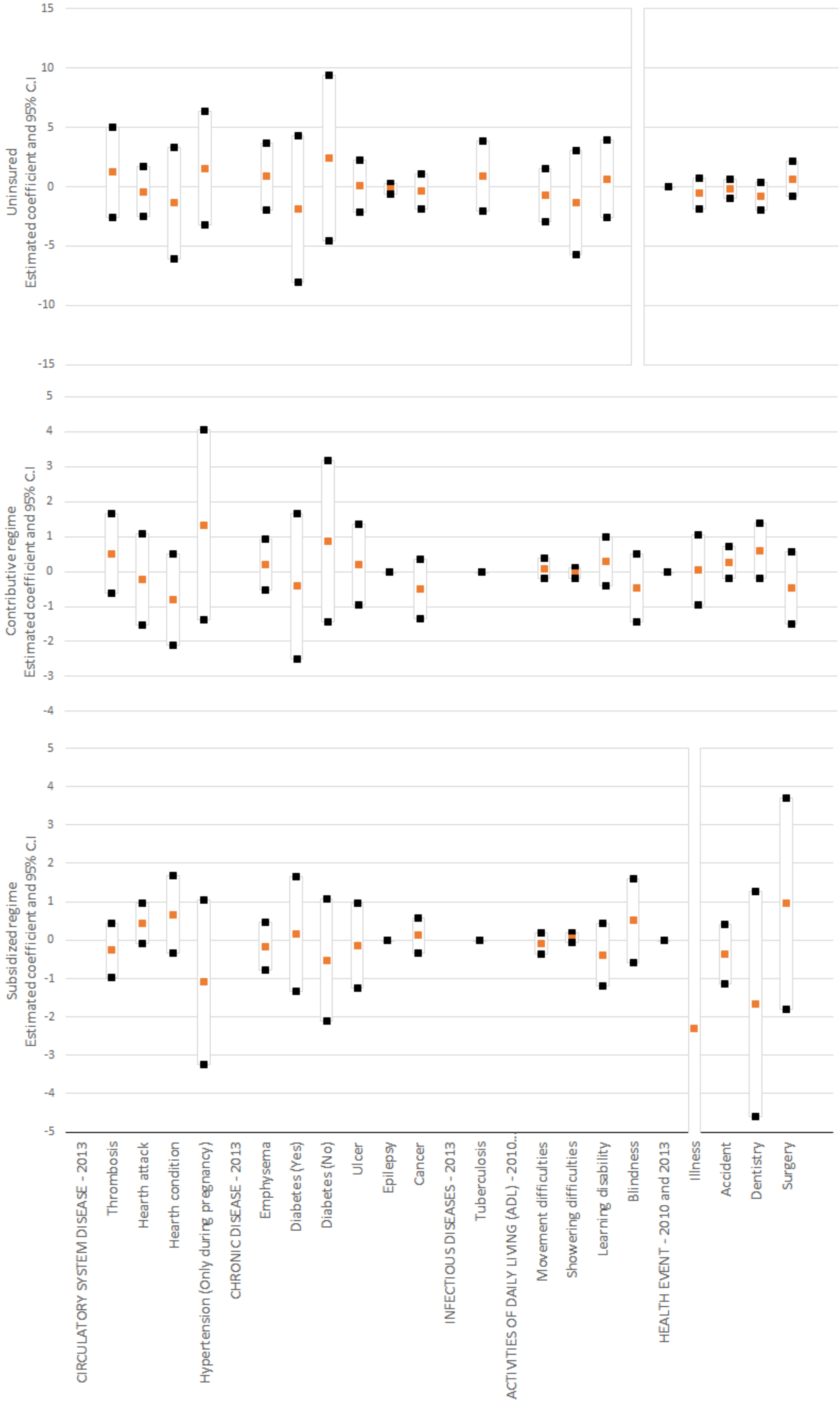
Regarding the interpretation of the coefficients, lets focus as an example in the effect of the subsidized regime on health status variables. In this case, the Subsidized Health regime has the strongest positive effect on the incidence of surgery, followed by hearth condition and blindness. On the other end, the Subsidized Health regime has the more negative strongest effect on illness, dentistry events and hypertension during pregnancy. Across health insurance status, being uninsured have the strongest effect across al variables when compared with individuals belonging to any of the two forms of insurance.

From this exercise it is clear that there are some variables for which the treatment does not present a significant effect due to high variability in the data, but for most of them the confidence intervals are small and therefore they estimate with precision the coefficients. This implies that either the effect of the treatments are truly zero or that the effect is very small.

The results on this paper contradicts the previous literature, as most of them find a statistically significant effect among SR and its respective control group in at least some response variables, which usually include some of the same variables in this study or a different proxy to measure the same latent variable. As explained before this may be due to the equalization and reaching universal health coverage in Colombia after 2009, as all of such studies use data before 2010.

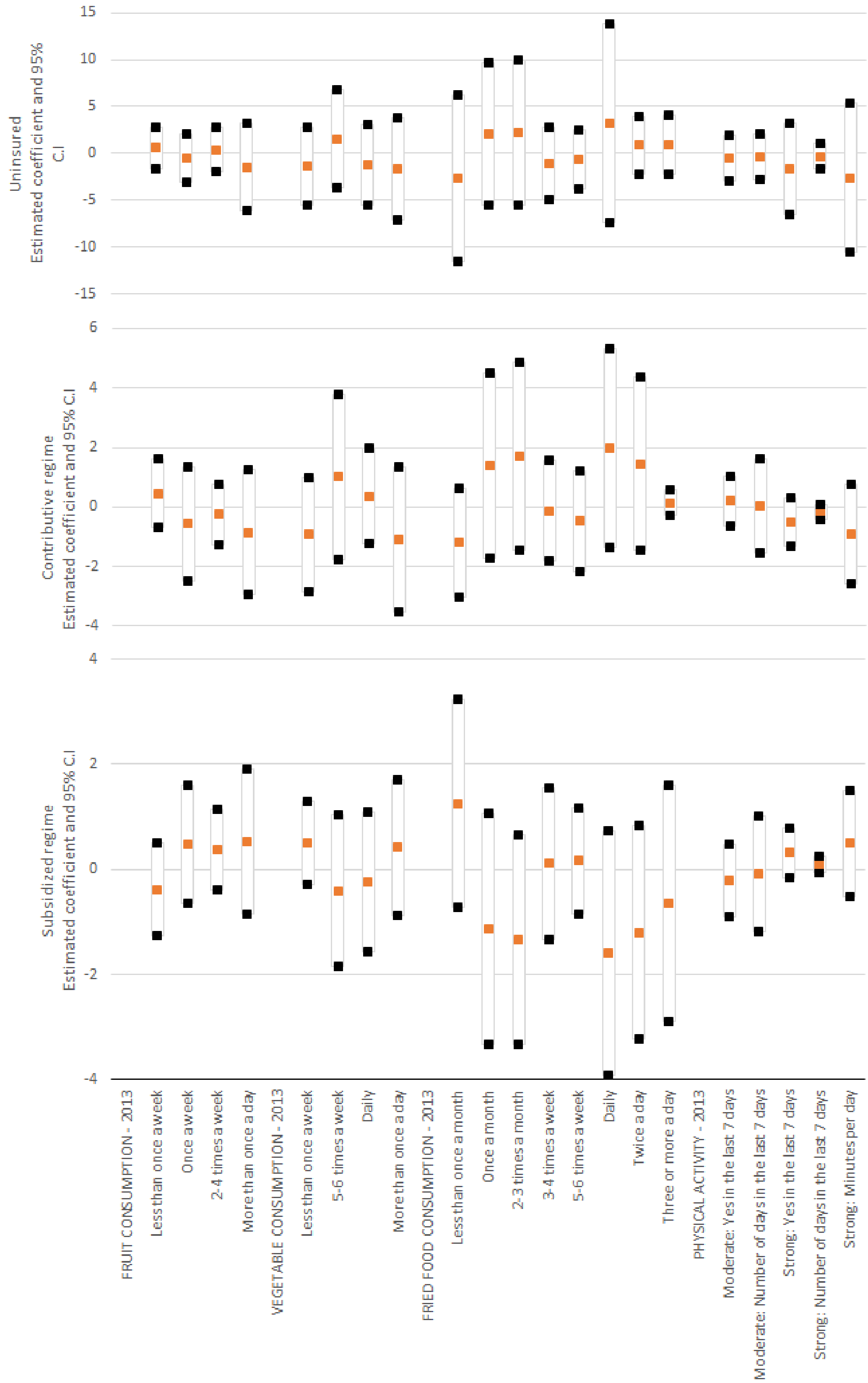
Finally, using standardized variables allows to compare the scale of the response of different endogenous variables to the treatment variable. On the one hand, the Subsidized Health Insurance Regime has the biggest positive impacts in the incidence of hypertension, diabetes during pregnancy, surgery, consumption of fried food less than once a month and visits to the doctor. On the other hand the same health insurance regime has the more negative impacts in incidence of illness, visits to the dentistry, daily consumption of fried food, hospitalizations in the last 12 months and the use of alternative medicines. In respect to the Contributive Health insurance status, the most positive effects are on the incidence of hypertension during pregnancy, asthma, an increase in the consumption of fried food 2-3 times per month and daily and an increase in the health annual expenditure. The more negative effects are on illness, dentistry events, consumption of vegetables in high frequency and a decrease in low frequency fried food consumption. Being uninsured present a positive effect in the incidence of hypertension during pregnancy (the highest among the insurance regimes), the highest incidence of asthma, highest vegetable consumption on a high frequency and highest consumption of fried food on low and high frequencies. However, none of the results are statistically significant as all of the confidence intervals contain the zero.

Figure 5: Estimated coefficient and 95% confidence interval: Health status variables.



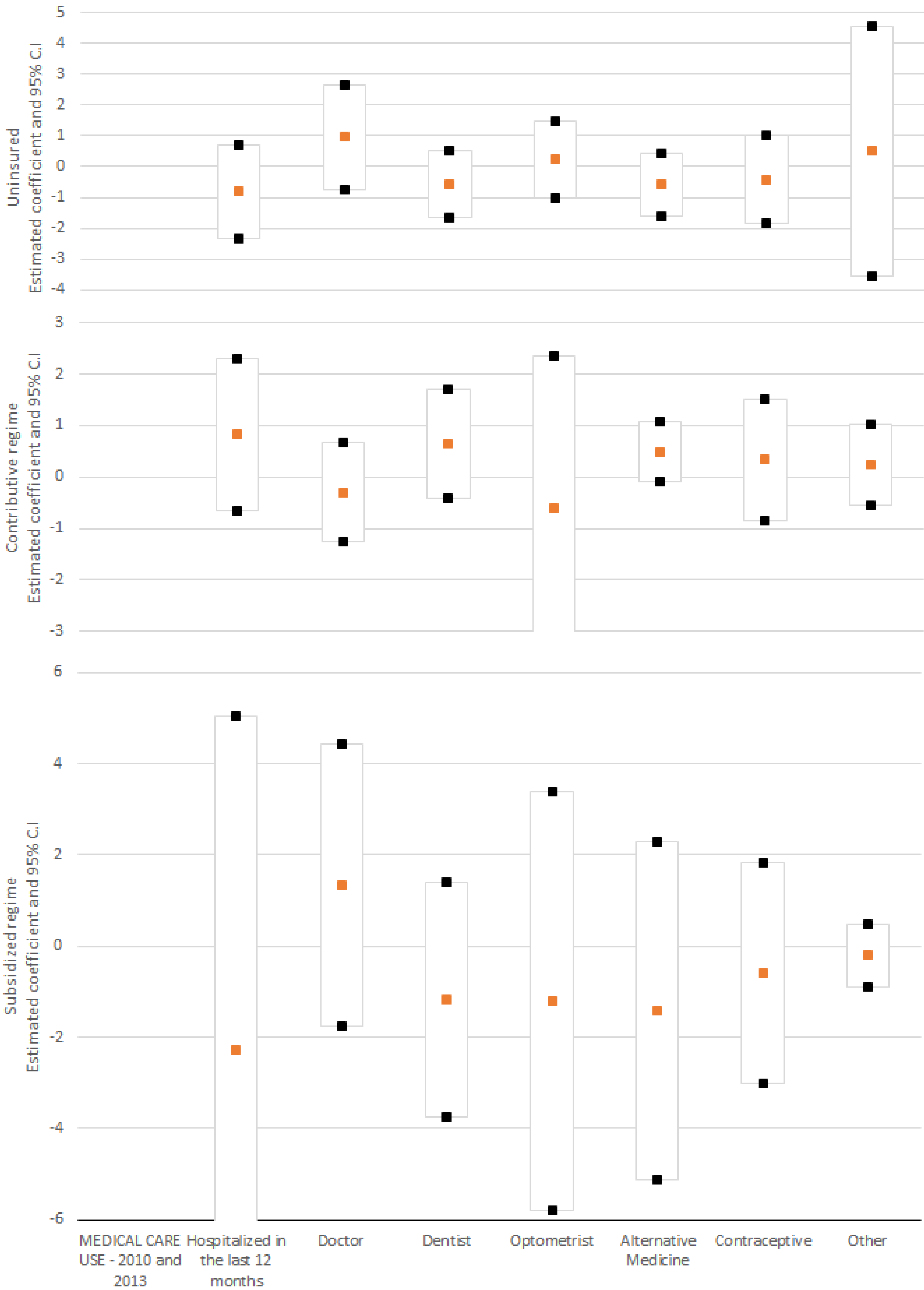
Source: Author's estimations using CLSA for the years 2013 and 2016 or 2010, 2013 and 2016 when indicated. Note: 1/ Conventional estimates with conventional standard errors. 2/ The order of the approximation of the polynomial for estimating the conventional coefficient is 1. 3/ Triangular kernel is used for the estimations and a data-driven bandwidth selection method. 4/ Standardized variables using $M_i^* = \frac{M_i - \overline{M_i}}{SD_{M_i}}$ where M_i corresponds to the response variable, the treatment variable and the running variable.

Figure 6: Estimated coefficient and 95% confidence interval: Behavioural distortion variables



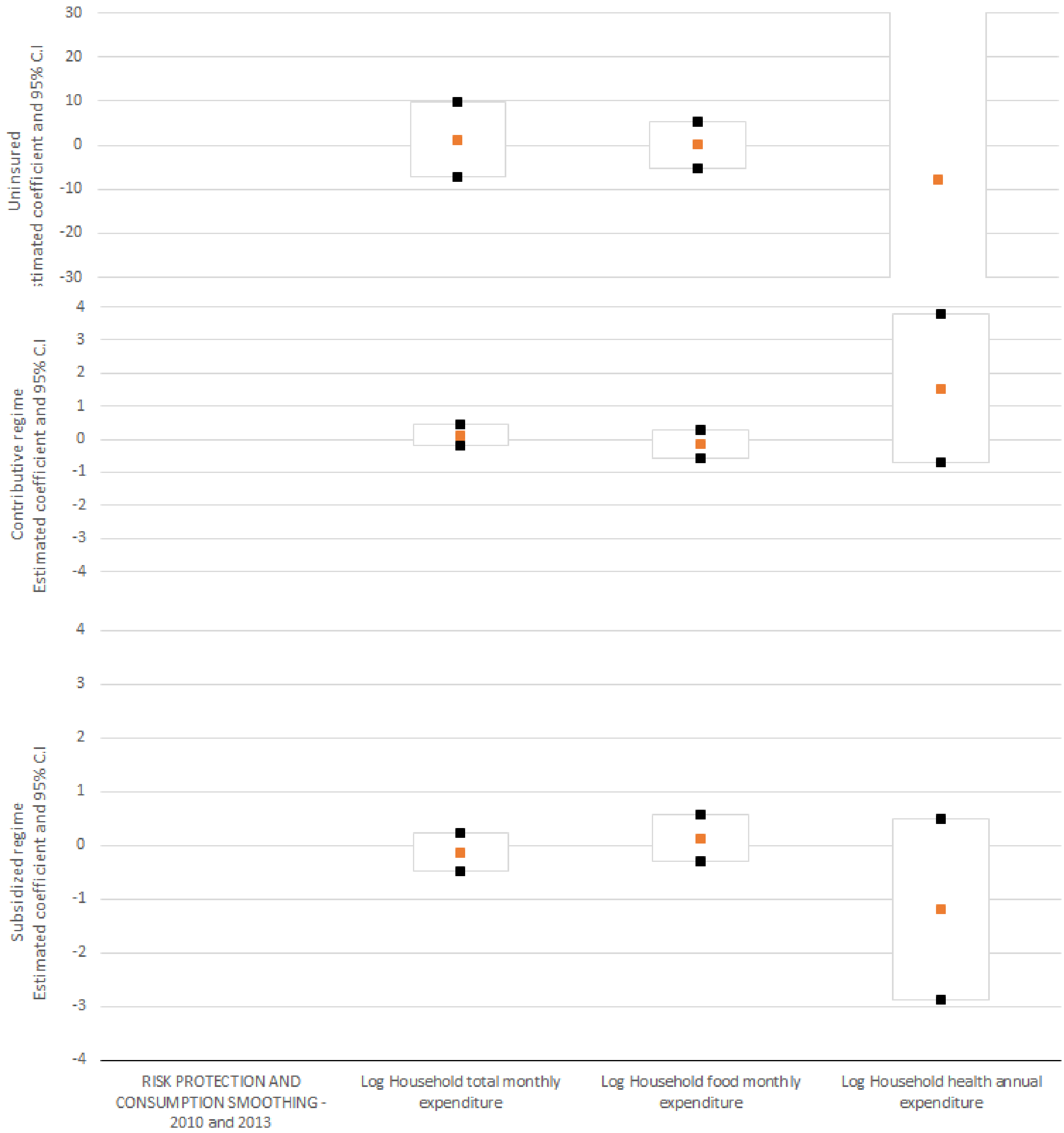
Source: Author's estimations using CLSA for the years 2013 or 2013 and 2016 when indicated.
Note: 1/ Conventional estimates with conventional standard errors. 2/ The order of the approximation of the polynomial for estimating the conventional coefficient is 1. 3/ Triangular kernel is used for the estimations and a data-driven bandwidth selection method. 4/ Standardized variables using $M_i^* = \frac{M_i - \overline{M_i}}{SD_{M_i}}$ where M_i corresponds to the response variable, the treatment variable and the running variable.

Figure 7: Estimated coefficient and 95% confidence interval: Medical care use variables



Source: Author's estimations using CLSA for the years 2010, 2013 and 2016 when indicated.
Note: 1/ Conventional estimates with conventional standard errors. 2/ The order of the approximation of the polynomial for estimating the conventional coefficient is 1. 3/ Triangular kernel is used for the estimations and a data-driven bandwidth selection method. 4/ Standardized variables using $M_i^* = \frac{M_i - \overline{M_i}}{SD_{M_i}}$ where M_i corresponds to the response variable, the treatment variable and the running variable.

Figure 8: Estimated coefficient and 95% confidence interval: Risk protection and Consumption smoothing variables



Source: Author's estimations using CLSA for the years 2010, 2013 and 2016 or 2013 and 2016 when indicated.
Note: 1/ Conventional estimates with conventional standard errors. 2/ The order of the approximation of the polynomial for estimating the conventional coefficient is 1. 3/ Triangular kernel is used for the estimations and a data-driven bandwidth selection method. 4/ Standardized variables using $M_i^* = \frac{M_i - \overline{M_i}}{SD_{M_i}}$ where M_i corresponds to the response variable, the treatment variable and the running variable.

5.4 Robustness checks

To verify the stability of the results numerous robustness checks were conducted. Besides trying different specification methods for the bandwidth selection as described previously, also different kernel functions were used to construct the local-polynomial estimator. The main results shown here used an Epanechnikov kernel, but the estimations were also made using Uniform and Triangular Kernel functions. In none of the cases the results change much. The local polynomial used to construct the point estimator in the main results shown here is of the first degree, but also other degrees of approximation were estimated. In particular second and fourth degree. The results show a statistically significant jump in the treatment variables when the linear approximation is used. Nevertheless, the existence of such a jump is not statistically clear when higher polynomial orders are used to approximate the fit of the first stage estimation.

One of the main concerns is that assuming the SISBEN score in 2016 being the same than in the last year reported before 2013 may induce measurement error in the running variable. To test the effect of this, the models were run only for the cohorts 2010 and 2013 and including only the persons who declare directly the insurance affiliation status and report their own response variables (head of the household and partner mainly). This greatly reduces the sample to 1573 observations for the variables with most observations and 554 for the variables with less observations. Although this, the lack of significance of the health insurance status on the response variables remains the same.

For the variables in the database which are available in the two initial years (2010 and 2013) -mainly some health status variables, medical care use variables, and expenditure variables-, the models were run for each year separately. The results in terms of significance and sign of the effects are the same in both years. Additionally, a method for correcting the p-values on the sense proposed by Bonferroni was applied. This is due to the fact that the estimations made here for the combination of the three possible treatment groups can be seen as sequentially applied using the same data, but changing the treatment and control status of the observations. In general, the result are the same. None of the treatments have any effect on the response variables, and such effect remains the same after applying the Benjamini and Liu, 1999a correction for simultaneous tests as shown in appendix 5A and 5B.

All of this set of results are available on request.

Finally, figure 3 indicates that when considering the three insurance status together, the jump in the affiliation to the Subsidized Health Regime is associated with a jump in the affiliation to the Contributive Health Regime given that the uninsured status present no jump at the eligibility threshold. This may be interpreted as evidence in favour of the hypothesis that all the identification in the model comes only from the type of insurance (Contributive vs Subsidized Health Regime) and not from the uninsured individuals. To test this, the Figure 3, and tables 3 to 6 were repeated but in this case dropping the uninsured individuals. Appendix 6 shows the discontinuity identification exercise, only among insured individuals. The jump in the affiliation status to the Subsidized Health Regime still exists, however it is smaller. While the jump in the affiliation status to the Subsidized Health regime at the eligibility threshold here is around 10%, it is of around 15% in figure 3. Furthermore, all the models were re-estimated ignoring the uninsured individuals. Appendix 7A to 7D show this results. Again, the results are the same: there is no difference in any of the response variables according to the type of insurance the individuals have.

Therefore a model considering only the Subsidized Health Regime and the Contributive Health Regime do not seem viable as it would ignore an important portion of the jump in the identification strategy and would generate selection bias in the sample. This work in particular follows the approximation of Camacho and Conover (2013) and Miller et al. (2013) preferring to have heterogeneous control groups instead of generating selection bias in the sample when ignoring the uninsured individuals.

6. Conclusions and outlook

This paper evaluates the effect of different health insurance types in Colombia over a wide set of variables grouped in health status, behavioural distortions, medical care use and risk protection and consumption smoothing. This is the first paper to model the existence of three health insurance statuses in Colombia: individuals belonging to the Subsidized Health Regime, individuals on the Contributive Health Regime and Uninsured Population.

For evaluating the particular effect among the health insurance type, I used data from the Colombian Longitudinal Survey by Universidad de los Andes for the years 2010, 2013 and 2016, while the econometric methodology used is a Fuzzy Regression Discontinuity Design.

The necessary assumptions were tested. In urban and rural areas there is minimal evidence of manipulation of the running variable, and there is balance in the pretreatment and exogenous variables among the treatment and control groups. Nevertheless, only in urban areas there is a jump in the treatment variable as function of the running variable, showing the appropriateness of the use of the Fuzzy Regression discontinuity design in urban areas. In the rural areas, there is no jump in the probability of receiving the treatment invalidating completely any analysis made trough Fuzzy Regression Discontinuity design in such areas. This implies that the results in this paper can only be extrapolated to the urban areas in Colombia.

Early works on the topic in urban Colombia, have found that there were significant differences between the Subsidized and Contributive Health Regimes, previous to the equalization of the Mandatory Health Plans in 2009, and, the reach of the universal healthcare coverage in the same year. This paper finds that after 2010 there seems to be no significant difference among the three possible health regimes: the Subsidized, the Contributive and the uninsured in the variables related to health status, medical care use, risk protection against illness and behavioural distortions. This may be due to the equalization of the Mandatory Health Plans in 2009 and the compulsory basic medical attention and emergency room services for uninsured population jointly with reaching universal coverage of the general health insurance system.

This result may also be due to an improvement on the Subsidized Health Regime, a deterioration of the Contributive Health Regime or a combination of both. The reason for this equalization is still to be explored. Nevertheless, not finding a significant difference among being uninsured and having any kind of insurance is a surprising result. This result may be due first, to not having effectively a different effect between having an insurance or not having it. Second, having a huge variation on the data which would lead to the imprecise measurement of the effect of the treatment on the response variables. Third, the effect may exist but be very small. In some cases, it was found that the number of observations for certain insurance groups was very small, leading to wide confidence intervals and very imprecise estimated effects, while in most of the cases the conclusion is that either, the effect of the treatment in the response variables is very small, or it is definitely not significant.

For mitigating the imprecision of the estimations a bigger sample is required. For this, it would be ideal to have access to the official registry of the SISBEN scores matched with data of use of healthcare services from the Health Ministry at national level.

7.

References

[1] Calonico, Sebastian, Matias Cattaneo and Rocio Titiunik. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*. 82(6): 2295-2326, 2014

[2] Calonico, Sebastian, Matias Cattaneo and Rocio Titiunik. Robust data-driven Inference in the Regression-discontinuity Design. *The Stata Journal*. 14(4): 909-946, 2014

[3] Calonico, Sebastian, Matias Cattaneo and Rocio Titiunik. rdrobust: Software for Regression-discontinuity Designs. *The Stata Journal*. 17(2): 372-404, 2017

[4] Camacho, Adriana, and Emily Conover. Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy*. 3(2): 41–65, 2011

[5] Camacho, Adriana and Emily Conover. Effects of Subsidized Health Insurance on Newborn Health in a Developing Country. *Economic Development and Cultural Change*. 61(3): 633-658, 2013

[6] Cattaneo, Matias, Michael Jansson, and Xinwei Ma. Simple Local Polynomial Density Estimators. Working paper, University of Michigan. 2017

[7] Cattaneo, Matias, Michael Jansson, and Xinwei Ma. Manipulation Testing based on Density Discontinuity. *Stata Journal*. 18(1): 234-261, 2018

[8] De la Mata, Dolores, and Carlos Felipe Gaviria. Losing health insurance when young: Impacts on usage of medical services and health. CAF – Working paper No. 2015/04, 2015

[9] Department of National Planning. Documento CONPES Social 117. Actualización de los criterios para la determinación e identificación y selección de beneficiarios de programas sociales. Bogotá, 2008

[10] Department of National Planning. Diseño del índice SISBEN en su tercera versión -SISBEN III-. Bogotá, 2008

[11] Econometria. Evaluacion de resultados e impacto de la unificacion del Plan Oblogatorio de Salud y una evaluacion de resultados de su mas reciente actualizacion, que permita identificar los avances en cuanto a equidad y acceso de las nuevas tecnologias en salud incluidas en los planes de beneficios. Bogotá, 2015

[12] Escobar, Maria Luisa, Ursula Giedion, Antonio Giuffrida and Amanda Glassman. Colombia: After a Decade of Health System Reform. In Glassman, A. *From Few to Many: Ten Years of Health Insurance Expansion in Colombia*. Inter-American Development Bank, Washington D.C, 1-13, 2009

[13] Enthoven, Alain C. The history and principles of managed competition. *Health Affairs*. 12, Suplement: 24–48. 1993

[14] Flórez Nieto CE, Camacho A. Dos décadas de cambios en la equidad del sistema de salud colombiano: 1990-2010. Santiago de Chile: CEPAL. 2012

[15] Gaviria, Alejandro, Carlos Medina, and Carolina Mejía. Assessing Health Reform in Colombia: From Theory to Practice. *Economía*. 7(1): 29–72, 2007

[16] Genoni, Maria. Health Shocks and Consumption Smoothing: Evidence from Indonesia. *Economic Development and Cultural Change*. 60(3): 475-506. 2012

[17] Gertler, Paul and Jonathan Gruber. Insuring Consumption against Illness. *American Economic Review*. 92(1): 51-70, 2002

[18] Giedion, Ursula, Beatrice Yadira Díaz, Educardo Andrés Alfonso, and William D. Savedoff. The Impact of Subsidized Health Insurance on Health Status and on Access to and Use of Health Services. In *From Few to Many: Ten Years of*

Health Insurance Expansion in Colombia, edited by Amanda L. Glassman, María-Luisa Escobar, Antonio Giuffrida, and Ursula Giedion, 47–74. Washington, DC: Inter-American Development Bank and The Brookings Institution. 2009

[19] Giedion, Ursula, and Manuela Villar Uribe. Colombia's Universal Health Insurance System. *Health Affairs*. 28(3): 853–863, 2009

[20] Houweling, Tanja AJ, Ivan Arroyave, Alex Burdorf, and Mauricio Avendano. Health insurance coverage, neonatal mortality and caesarean section deliveries: an analysis of vital registration data in Colombia. *Journal of Epidemiology and Community Health*. 71(5): 505-512, 2017

[21] Imbens, Guido W. and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*. 142(2): 615-635, 2008

[22] International Labour Office (ILO), Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), and World Health Organization (WHO). Berlin Recommendations for Action. In *The International Conference on Social Health Insurance in Developing Countries*. Berlin, 05–07 December 2005. Geneva: ILO, GTZ, and WHO. 2006

[23] Lee, David and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of economic literature*. 48: 281-355, 2010

[24] Levy, Helen. The Economic Consecuences of Being Uninsured. Working Paper Series. Economic Research Initiative on the Uninsured. 2002

[25] Londoño Juan, and J. Frenk. Structured Pluralism: Towards an Innovative Model for Health System Reform in Latin America. *Health Policy*. 41:1–36. 1997

[26] Miller, Grant, Diana Pinto, and Marcos Vera-Hernández. Risk Protection, Service Use, and Health Outcomes under Colombia's Health Insurance Program for the Poor. *American Economic Journal: Applied Economics*. 5(4): 61-91, 2013

[27] Morduch, Jonathan. Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives*. 9(3): 103–14. 1995

[28] Ministerio de Salud y Protección Social. Consejo Nacional de Seguridad Social en Salud. Acuerdo 260 de 2004.

[29] Ministerio de Salud y Protección Social. Sistema de Salud en Colombia. In Instituto Suramericano de Gobierno en Salud. *Sistemas de Salud en Suramerica: desafíos para la universalidad, la integralidad y la equidad*. Rio de Janeiro, 349-440, 2012

[30] Ministerio de Salud y Protección Social.¿Cómo operan las cuotas de recuperación y la tarifa plena?. *www.minsalud.gov.co*. 14-Nov-2013. Web. 14-Aug-2017

[31] Ministerio de Salud y Protección Social.Estudio de sostenibilidad del aseguramiento en salud para el año 2017. *www.minsalud.gov.co*. Web. 09-May-2019

[32] Mohanan, Manoj. Causal Effects of Health Shocks on Consumption and Debt: Quasi-Experimental Evidence from Bus Accident Injuries. *Review of Economic and Statistics*. 95(2): 673–81. 2013

[33] Oldenburg, Catherine, Ellen Moscoe and Till Bärnighausen. Regression Discontinuity for Causal Effect Estimation in Epidemiology. *Current Epidemiology Reports*. 3: 233–241, 2016

[34] Paxson, Christina. Consumption and Income Seasonality in Thailand. *Journal of Political Economy*. 101(1): 39–72. 1993

[35] Prada, Sergio, Jorge Duarte, and Ramiro Guerrero. Out-of-pocket health expenditure for poor and non-poor older adults in Colombia: composition and trends. *International journal of consumer studies*. 39(4): 362-368, 2015

[36] Sandoval, Carlos. Welfare and poverty traps in urban Colombia: an approach from the asset dynamics perspective. Working paper. 2019

[37] Santa María, Mauricio, Fabián García, Carlos F. Prada, María J. Uribe, and Tatiana Vásquez. El sector salud en Colombia: impacto del SGSSS después de más de una década de la reforma. *Coyuntura Social*. No. 38. Fedesarrollo, June 2008

[38] Smith, James. Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status. *Journal of Economic Perspectives*. 13(2): 145-166, 1999

[39] Strauss, John, Paul Gertler, Omar Rahman and Kristin Fox. Gender and Life-Cycle Differentials in the Patterns and Determinants of Adult Health. *The Journal of Human Resources*. 28(4): 791-837. 1993

[40] Strauss, John, and Duncan Thomas. Health, Nutrition, And Economic Development. *Journal of Economic Literature*. 36(2): 766-817. 1998

[41] Thomas, Duncan and Elisabeth Frankenberg. The Measurement and Interpretation of Health in Social Surveys. RAND. Working paper series 01-06. 2000

[42] Townsend, Robert. Risk and Insurance in Village India. *Econometrica* 62(3): 539–91. 1994

[43] Trujillo, Antonio, Jorge Portillo, and John Vernon. The Impact of Subsidized Health Insurance for the Poor: Evaluating the Colombian Experience Using Propensity Score Matching. *International Journal of Health Care Finance and Economics*. 5(3): 211–39. 2005

[44] Universidad de los Andes. Colombia en Movimiento. Un análisis descriptivo basado en la Encuesta Longitudinal Colombiana de la Universidad de los Andes ELCA. Universidad de los Andes, CEDE. 2011

[45] Universidad de los Andes. Colombia en Movimiento 2010-2013. Los cambios en la vida de los Hogares a través de la Encuesta Longitudinal Colombiana de la Universidad de los Andes ELCA. Universidad de los Andes, CEDE. 2014

[46] Universidad de los Andes. Colombia en Movimiento 2010-2013-2016. Los cambios en la vida de los Hogares a través de la Encuesta Longitudinal Colombiana de la Universidad de los Andes (ELCA). Universidad de los Andes, CEDE. 2017

[47] Wagstaff, Adam. The Economic Consequences of Health Shocks: Evidence from Vietnam. *Journal of Health Economics* 26(1): 82-100. 2007

[48] World Health Organization and International Bank for Reconstruction and Development.Tracking universal health coverage: 2017 global monitoring report. The World Bank, Licence: CC BY-NC-SA 3.0 IGO. 2017.

8. Appendix

Appendix 1A: Health Status Variables. Variable definition, survey question and target population.

Econometric model variable	Survey question	Cohort available 2010	2013	2016	Target population	Zone
CIRCULATORY SYSTEM DISEASE						
Thrombosis	Have you ever had thrombosis or a brain stroke?		X	X	For 2013: Head of the household and partner in 2013. For 2016: Head of the household and partner selected in 2010 as part of the panel.	Urban and Rural
Hearth attack	Have you ever had a hearth attack?		X	X		
Hearth condition	Have you ever been diagnosed by a doctor or healthcare professional with: A hearth condition		X	X		
Hypertension (Yes)	Have you ever been diagnosed by a doctor or healthcare professional with: Hypertension?		X	X		
Hypertension (Only during pregnancy)	Have you ever been diagnosed by a doctor or healthcare professional with: Hypertension, only during pregnancy?		X	X		
CHRONIC DISEASE						
Asthma	Have you ever been diagnosed by a doctor or healthcare professional with: Asthma?		X	X		
Emphysema	Have you ever been diagnosed by a doctor or healthcare professional with: any long-term respiratory disease, like emphysema or chronic bronchitis?		X	X		
Diabetes (Yes)	Have you ever been diagnosed by a doctor or healthcare professional with: diabetes or high blood sugar?		X	X		
Diabetes (No)	Have you ever been diagnosed by a doctor or healthcare professional with: diabetes or high blood sugar?		X	X		
Diabetes (Only during pregnancy)	Have you ever been diagnosed by a doctor or healthcare professional with: diabetes or high blood sugar, only during pregnancy?		X	X		
Ulcer	Have you ever been diagnosed by a doctor or healthcare professional with: stomach or intestinal ulcer?		X	X		
Epilepsy	Have you ever been diagnosed by a doctor or healthcare professional with: epilepsy or seizures?		X	X		
Cancer	Have you ever been diagnosed by a doctor or healthcare professional with: cancer?		X	X		
INFECTIOUS DISEASES						
Tuberculosis	Have you ever been diagnosed by a doctor or healthcare professional with: Tuberculosis?		X	X		
ACTIVITIES OF DAILY LIVING (ADL)						
Movement difficulties	Due to an accident, illness or a birth condition do you have any of these conditions permanently?. Movement difficulty or inability to walk by yourself	X	X	X	For 2010: children between 0 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	Urban and Rural
Showering difficulties	Due to an accident, illness or a birth condition do you have any of these conditions permanently?. Showering difficulty, dressing difficulty or inability to feed by yourself	X	X	X		
Learning disability	Due to an accident, illness or a birth condition do you have any of these conditions permanently?. Difficulties to learn and understand	X	X	X		
Blindness	Due to an accident, illness or a birth condition do you have any of these conditions permanently?. Blindness	X	X	X		
HEALTH EVENT						
Illness	In the last 30 days, have you had any of these health problems that did not imply hospitalization?. Illness or chronic pain	X	X	X	For 2010: Head of the household and partner. Children between 0 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel.	Urban and Rural
Accident	In the last 30 days, have you had any of these health problems that did not imply hospitalization?. Accident or lesion	X	X	X		
Dentistry	In the last 30 days, have you had any of these health problems that did not imply hospitalization?. Dental problem	X	X	X		
Surgery	In the last 30 days, have you had any of these health problems that did not imply hospitalization?. Ambulatory surgery	X	X	X		
Pregnancy	In the last 30 days, have you had any of these health problems that did not imply hospitalization?. Complications of pregnancy, postdelivery or abortion		X	X	Women only. For 2013: Head of the household or partner in 2013. For 2016: Head of the household and partner selected in 2010 as part of the panel.	Urban and Rural

Source: CLSA questionnaires for the years 2010, 2013 and 2016.

Appendix 1B: Behavioural Distortion Variables. Variable definition, survey question and target population.

Econometric model variable	Survey question	Cohort available 2010	2013	2016	Target population	Zone
FRUIT CONSUMPTION						
Less than once a week	How frequently do you eat fruit per week (as a whole or in juice)? (Including orange, apple, watermelon, guava, etc). Less than once a week		X	X	For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel.	Urban and Rural
Once a week	How frequently do you eat fruit per week (as a whole or in juice)? (Including orange, apple, watermelon, guava, etc). Once a week		X	X		
2-4 times a week	How frequently do you eat fruit per week (as a whole or in juice)? (Including orange, apple, watermelon, guava, etc). 2-4 times a week		X	X		
5-6 times a week	How frequently do you eat fruit per week (as a whole or in juice)? (Including orange, apple, watermelon, guava, etc). 5-6 times a week		X	X		
Daily	How frequently do you eat fruit per week (as a whole or in juice)? (Including orange, apple, watermelon, guava, etc). Daily, everyday		X	X		
More than once a day	How frequently do you eat fruit per week (as a whole or in juice)? (Including orange, apple, watermelon, guava, etc). Everyday, more than once a day		X	X		
VEGETABLE CONSUMPTION						
Less than once a week	How frequently do you eat vegetables per week (raw, cooked or in a soup)? (Including spinach, carrot, lettuce, chard, eggplant, artichoke etc). Less than once a week		X	X		
2-4 times a week	How frequently do you eat vegetables per week (raw, cooked or in a soup)? (Including spinach, carrot, lettuce, chard, eggplant, artichoke etc). 2-4 times a week		X	X		
5-6 times a week	How frequently do you eat vegetables per week (raw, cooked or in a soup)? (Including spinach, carrot, lettuce, chard, eggplant, artichoke etc). 5-6 times a week		X	X		
Daily	How frequently do you eat vegetables per week (raw, cooked or in a soup)? (Including spinach, carrot, lettuce, chard, eggplant, artichoke etc). Daily, everyday		X	X		
More than once a day	How frequently do you eat vegetables per week (raw, cooked or in a soup)? (Including spinach, carrot, lettuce, chard, eggplant, artichoke etc). Everyday, more than once a day		X	X		
FRIED FOOD CONSUMPTION						
Less than once a month	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. Less than once a month		X	X	For 2013: Head of the household and partner in 2013. For 2016: Head of the household and partner selected in 2010 as part of the panel.	Urban and Rural
Once a month	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. Once a month		X	X		
2-3 times a month	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. 2-3 times a month		X	X		
Twice a week	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. Twice a week		X	X		
3-4 times a week	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. 3-4 times a week		X	X		
5-6 times a week	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. 5-6 times a week		X	X		
Daily	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. Daily, once a day		X	X		
Twice a day	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. Twice a day		X	X		
Three or more a day	Usually, how frequently do you consume fried food per month (fries, fried meet, fried plantain, etc)?. Three or more a day		X	X		
PHYSICAL ACTIVITY						
Moderate: Yes in the last 7 days	In the last seven days have you done any moderate physical activity?. Yes		X	X	For 2013: Head of the household and partner in 2013. For 2016: Head of the household and partner selected in 2010 as part of the panel.	Urban and Rural
Moderate: Number of days in the last 7 days	In the last seven days have you done any moderate physical activity?. How many days?		X	X		
Moderate: Minutes per day	In the last seven days have you done any moderate physical activity?. How many minutes per day?		X	X		
Strong: Yes in the last 7 days	In the last seven days have you done any strong physical activity?. Yes		X	X		
Strong: Number of days in the last 7 days	In the last seven days have you done any strong physical activity?. How many days?		X	X		
Strong: Minutes per day	In the last seven days have you done any strong physical activity?. How many minutes per day?		X	X		

Source: CLSA questionnaires for the years 2010, 2013 and 2016.

Appendix 1C: Medical Care Use, Risk Protection and Running Variables. Variable definition, survey question and target population.

Econometric model variable	Survey question	Cohort available 2010	2013	2016	Target population	Zone
MEDICAL CARE USE						
Hospitalized in the last 12 months	Have you been hospitalized in the last 12 months?	X	X	X	For 2010: Head of the household and partner. Children between 0 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel.	Urban and Rural
Doctor	Without been sick, and only as preventive measure do you visit a general practitioner or medical specialist (gynecologist, urologist, cardiologist) at least once a year?	X	X	X		
Dentist	Without been sick, and only as preventive measure do you visit a dentist at least once a year?	X	X	X		
Optometrist	Without been sick, and only as preventive measure do you visit an optometrist at least once a year?	X	X	X		
Alternative Medicine	Without been sick, and only as preventive measure do you visit a professional in alternative medicine (homeopath or acupuncturist) at least once a year?	X	X	X		
Contraceptive	Without been sick, and only as preventive measure do you visit any family planning service at least once a year?	X	X	X		
Other	Without been sick, and only as preventive measure do you visit any other healthcare professional at least once a year?		X	X		
RISK PROTECTION AND CONSUMPTION SMOOTHING						
Log Household total monthly expenditure	In total, how much is the monthly expenditure of this household?	X	X	X	For 2010, 2013 and 2016: Household	Urban and Rural
Log Household food monthly expenditure	In total, how much is the monthly expenditure on food of this household?	X	X	X		
Log Household health annual expenditure	In the last 12 months, any member of the household have spent in healthcare?. How much?		X	X		
RUNNING VARIABLE						
SISBEN Score	Have this household ever been interviewed by the SISBEN?. Yes		X		For 2013: Household	Urban and Rural
	What was the last year when this household took the SISBEN survey?		X			
	Which was the score and level that the household reached in the last SISBEN survey? (Score goes from 0-100 and level from 1-6)		X			

Source: CLSA questionnaires for the years 2010, 2013 and 2016.

Appendix 1D: Affiliation Status Variables. Variable definition, survey question and target population.

Econometric model variable	Survey question	Cohort available 2010	2013	2016	Target population	Zone
AFFILIATION STATUS						
Affiliated to the subsidized health regime	Are you currently affiliated, paying by yourself or are you a beneficiary of any health insurance company?. Yes	X	X	X	For 2010: Head of the household and partner. Children between 0 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	Urban and Rural
	You are affiliated or covered by a health insurance because: SISBEN score allows you		X	X	For 2013: Head of the household and partner in 2013. For 2016: Head of the household and partner selected in 2010 as part of the panel.	
	To which of the following health insurance regimes are you affiliated?. Subsidized	X		X	For 2010: Head of the household and partner. For 2016: Persons 0 to 64 years old that are not part of the panel.	
	The child is affiliated or covered by a health insurance because: Is benefficiary of some household member. Whom?		X	X	For 2013: Children from 0 to 13 years old belonging to the panel. For 2016: Children from 6 to 16 years old belonging to the panel. Impute as health insurance regime for the kid, the health insurance regime declared by the adult in charge.	
	The child is affiliated or covered by a health insurance because: Is affiliated to the subsidized regime		X	X	For 2013: Children from 0 to 13 years old belonging to the panel. For 2016: Children from 6 to 16 years old belonging to the panel.	
Affiliated to the contributive health regime	Are you currently affiliated, paying by yourself or are you a beneficiary of any health insurance company?. Yes	X	X	X	For 2010: Head of the household and partner. Children between 0 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	
	You are affiliated or covered by a health insurance because: You are affiliated to any special regime (ARMY, Police Force, ECOPETROL, Public University or Public School Teachers)		X	X	For 2013: Head of the household and partner in 2013. For 2016: Head of the household and partner selected in 2010 as part of the panel.	
	You are affiliated or covered by a health insurance because: discount from payroll or you pay by yourself the public health insurance.		X	X		
	You are affiliated or covered by a health insurance because: discount from payroll or you pay by yourself other health insurance.		X	X		
	You are affiliated or covered by a health insurance because: you are beneficiary of an employee or a pensioner who is affiliated to a health insurance.		X	X		
	In your current job do you have…?. Health Insurance		X	X		
	In your current job do you have…?. Private health insurance paid by your employer		X	X		
	To which of the following health insurance regimes are you affiliated?. Contributive (EPS)	X		X	For 2010: Head of the household and partner. For 2016: Persons 0 to 64 years old that are not part of the panel.	
	The child is affiliated or covered by a health insurance because: Is benefficiary of some household member. Whom?		X	X	For 2013: Children from 0 to 13 years old belonging to the panel. For 2016: Children from 6 to 16 years old belonging to the panel. Impute as health insurance regime for the kid, the health insurance regime declared by the adult in charge.	
Uninsured	Are you currently affiliated, paying by yourself or are you a beneficiary of any health insurance company?. No	X	X	X	For 2010: Head of the household and partner. Children between 0 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	

Source: CLSA questionnaires for the years 2010, 2013 and 2016.

Appendix 1E: Exogenous Variables I. Variable definition, survey question and target population.

Econometric model variable	Survey question	Cohort 2010	Cohort 2013	Cohort 2016	Target population	Zone
EXOGENOUS VARIABLES (I)						Urban and Rural
Age	How old are you?	X	X	X	For 2010 and 2013: All household members. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. New household members.	
Proportion of household members in a given age group	How old are you?	X	X	X	For 2010 and 2013: All household members. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. New household members.	
	Household identifier	X	X	X	For 2010, 2013 and 2016: Household	
Years of education	What is the highest educational level and last grade approved by you on said level?	X	X	X	For 2010: Head of the household and partner. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. For 2016:Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	
	Are you currently studying (attending school or college)?	X	X	X	For 2010: Head of the household and partner. Children between 5 and 9 years old. For 2013: Head of the household and partner in 2013. Children from 0 to 13 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	
	What is your main occupation? To study	X	X	X	For 2010: Inactive head of the household and/or partner. For 2013: Inactive head of the household and/or partner in 2013. Inactive persons 0 to 64 years old that are not part of the panel. For 2016: Inactive head of the household and/or partner selected in 2010 as part of the panel. Inactive persons 0 to 64 years old that are not part of the panel.	
Per-capita income divided by poverty line	Usually what is the monthly income of this household due to: working or employment Usually what is the monthly income of this household due to: pensions Usually what is the monthly income of this household due to: real estate leasing Usually what is the monthly income of this household due to: interests and profits Usually what is the monthly income of this household due to: financial aid Usually what is the monthly income of this household due to: other income different from remittances	X	X	X	For 2010, 2013 and 2016: Household	
	In the last 12 months, did any member of this household receive income due to: inheritance	X	X	X		
	In the last 12 months, did any member of this household receive income due to: windfall		X	X		
	In the last 12 months, did any member of this household receive income due to: insurance policies	X	X	X		
	In the last 12 months, did any member of this household receive income due to: selling real state	X	X	X		
	In the last 12 months, did any member of this household receive income due to: selling a business	X	X	X		
	In the last 12 months, did any member of this household receive income due to: selling other assets	X	X	X		
	In the last 12 months, did any member of this household receive income due to: other income. Which one?	X	X	X		
	Household identifier	X	X	X	For 2010, 2013 and 2016: Household	
	Poverty lines	X	X	X	By the Colombian National Statistics Department (DANE)	

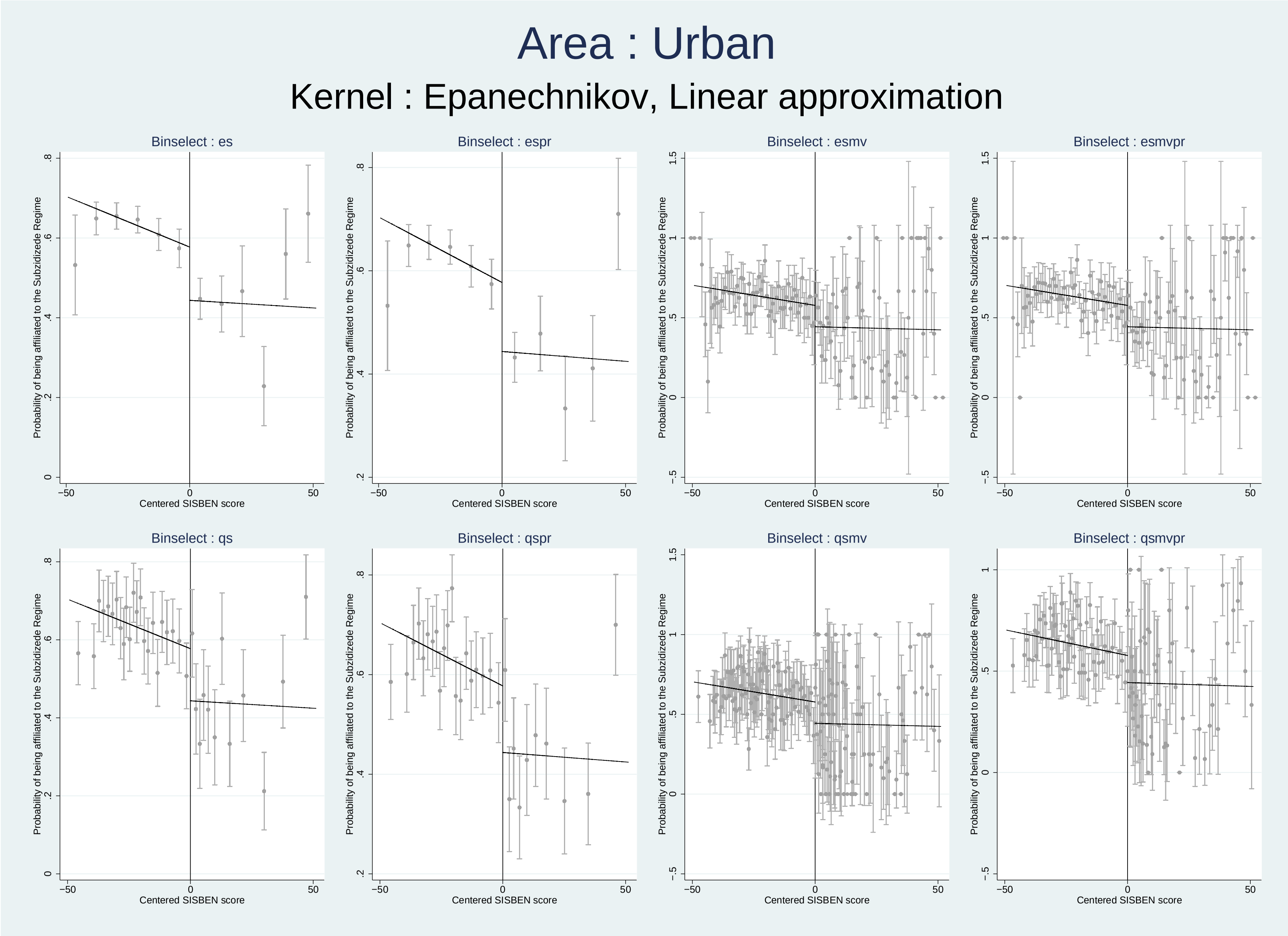
Source: CLSA questionnaires for the years 2010, 2013 and 2016.

Appendix 1F: Exogenous Variables II. Variable definition, survey question and target population.

Econometric model variable	Survey question	Cohort available 2010	2013	2016	Target population	Zone
EXOGENOUS VARIABLES (II)						
Gender: Male	Gender: Male	X	X	X	For 2010 and 2013: All household members. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children from 6 to 16 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. New household members.	Urban and Rural
Household size	Household identifier	X	X	X	For 2010, 2013 and 2016: Household	
Ocupation status: wage earner	In this job you are: A wage earner employee from a private firm with a long-term contract In this job you are: A wage earner employee from a private firm with a fix-term contract In this job you are: A wage earner employee from a public institution with a long-term contract In this job you are: A wage earner employee from a public institution with a fix-term contract	X	X	X	For 2010: Head of the household and partner. For 2013: Head of the household and partner in 2013. Children older than 10 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children older than 10 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	
Ocupation status: self-employed	In this job you are: pawn In this job you are: paid house worker In this job you are: self-employed In this job you are: employer In this job you are: worker in a farm (your own farm or leased)					
Ocupation status: no income earner	In this job you are: non-paid family worker					
Economic Status: Employee	During the last week, did you work at least one hour in a paid job? During the last week, did you work at least one hour as a non-paid family worker? During the last week, you didn't work, but you had a job of at least one hour? During the last week, did you work at least one hour and were you looking for a job?	X	X	X	For 2010: Head of the household and partner. For 2013: Head of the household and partner in 2013. Children older than 10 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel. For 2016: Head of the household and partner selected in 2010 as part of the panel. Children older than 10 years old belonging to the panel. Persons 0 to 64 years old that are not part of the panel.	
Economic Status: Inactive	During the last week, are you permanently handicapped for working?					
Economic Status: Unemployed	During the last week, none of the previous options?... Jump to the questions for unemployed population					

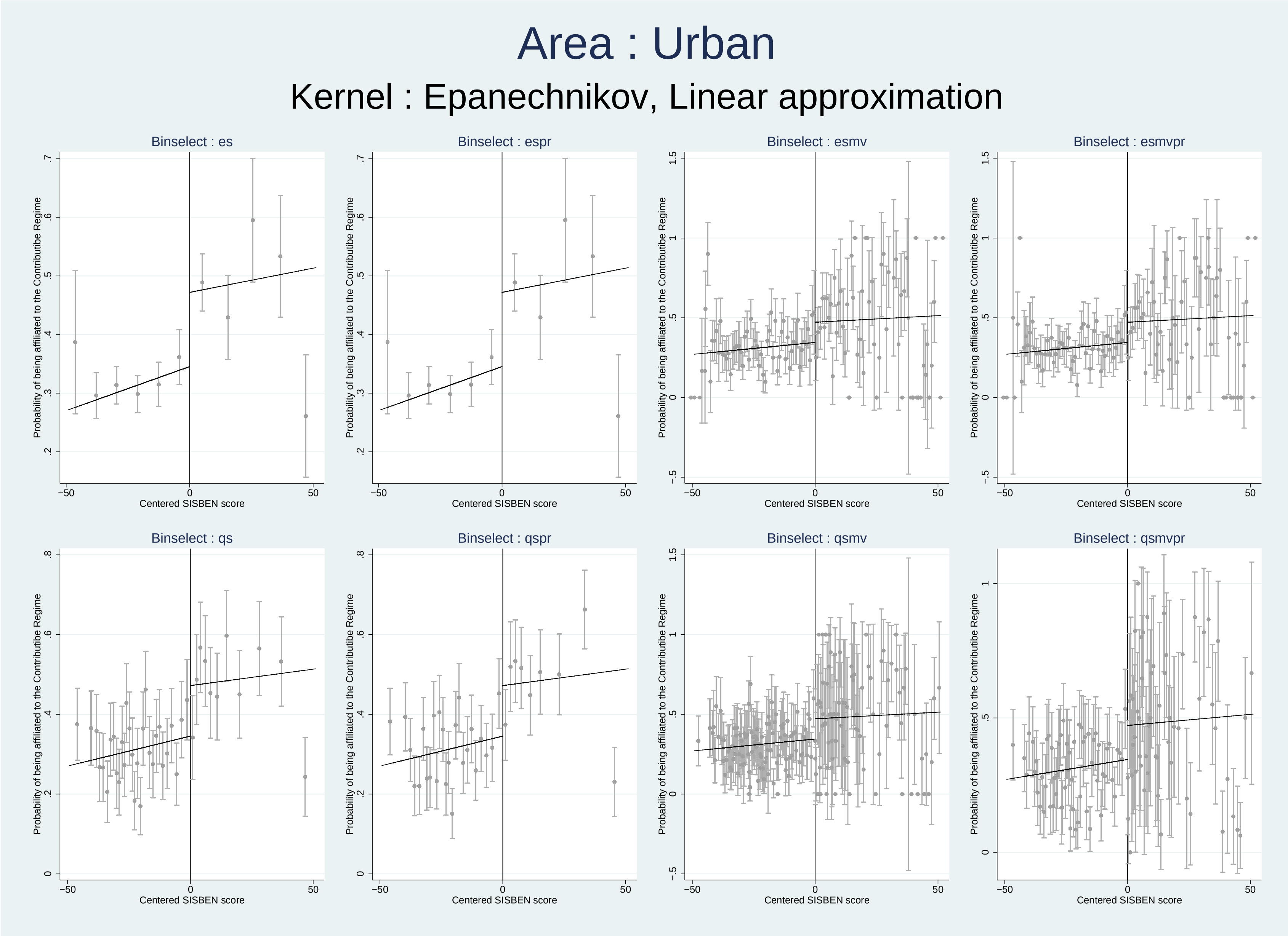
Source: CLSA questionnaires for the years 2010, 2013 and 2016.

Appendix 2A: Discontinuity identification: Subsidized Regime. Different Bandwidth Select Methods (I)



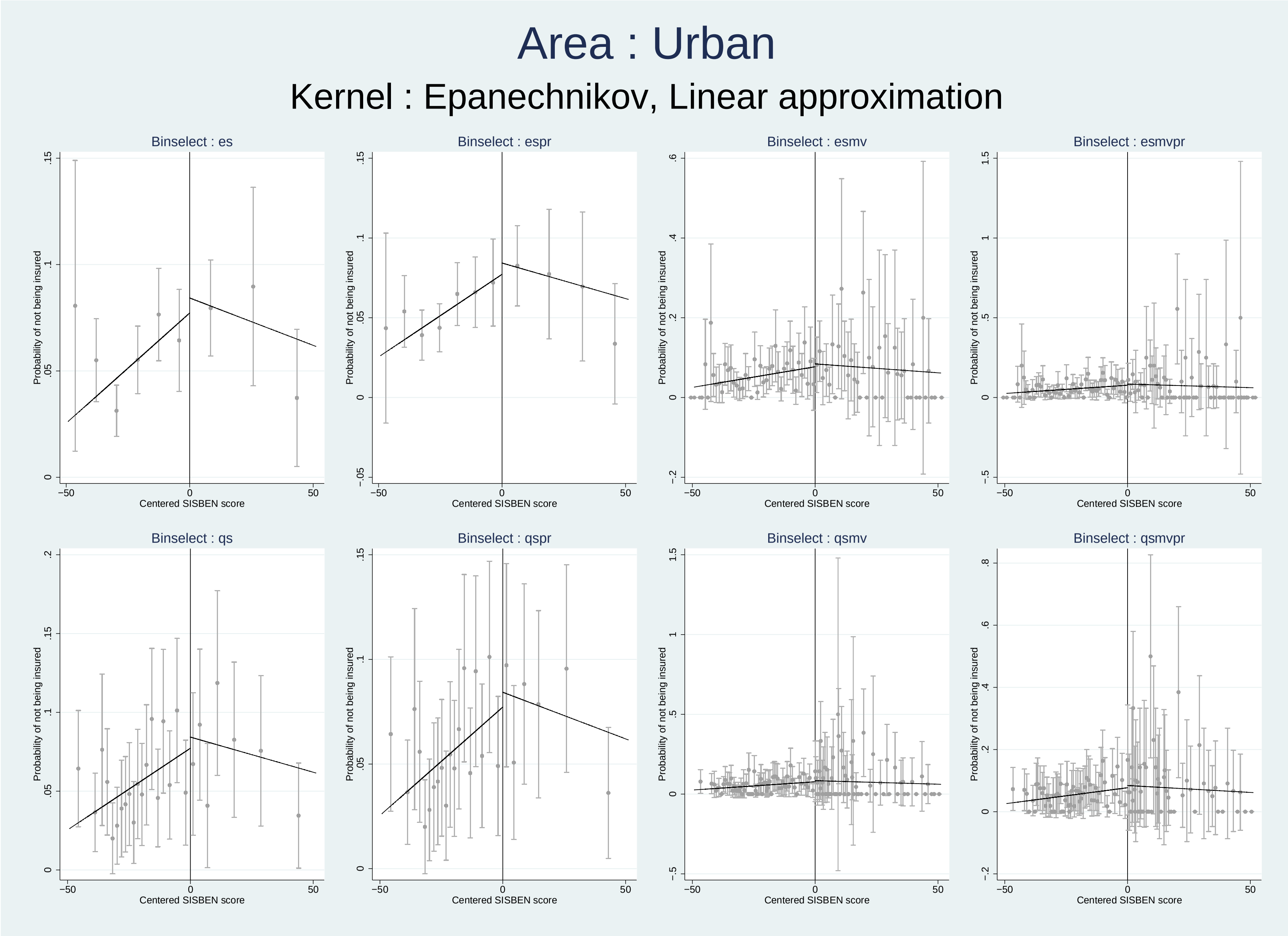
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the subsidized health insurance status. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: All methods available in STATA

Appendix 2B: Discontinuity identification: Contributive Regime. Different Bandwidth Select Methods (II)



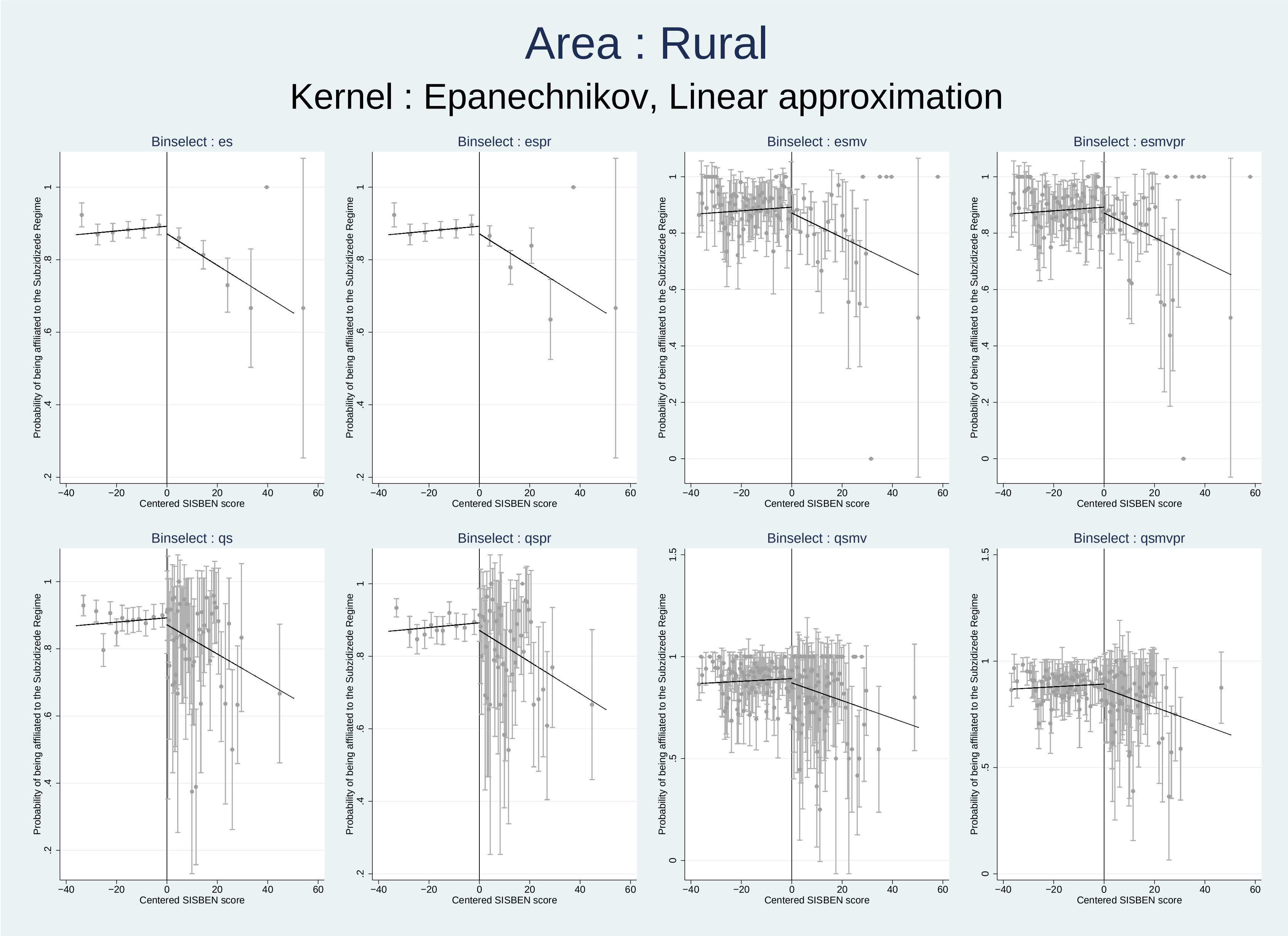
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the contributive health insurance status. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: All methods available in STATA

Appendix 2C: Discontinuity identification: Uninsured. Different Bandwidth Select Methods (III)



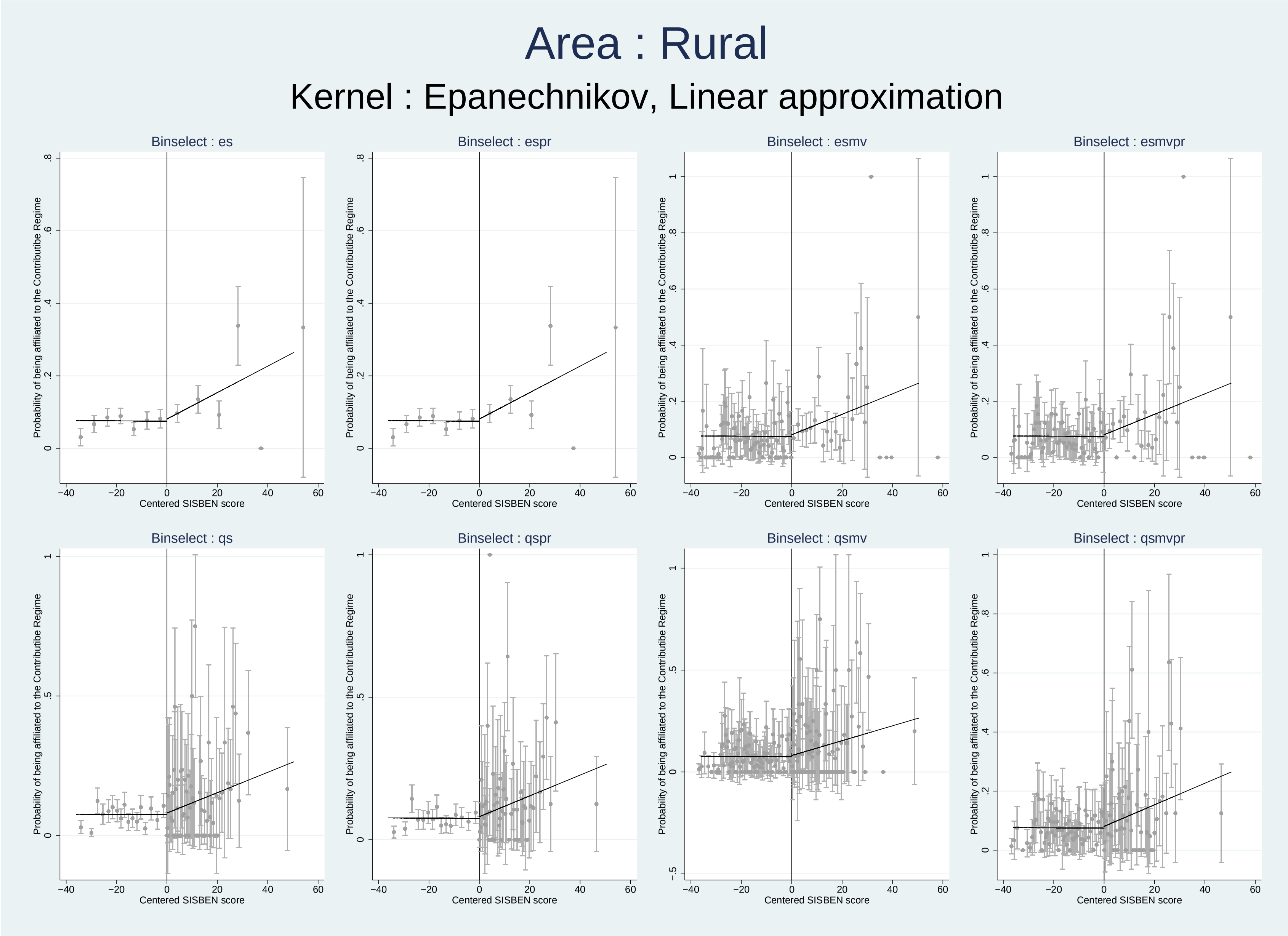
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the uninsured individuals. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: All methods available in STATA

Appendix 2D: Discontinuity identification: Subsidized Regime. Different Bandwidth Select Methods (IV)



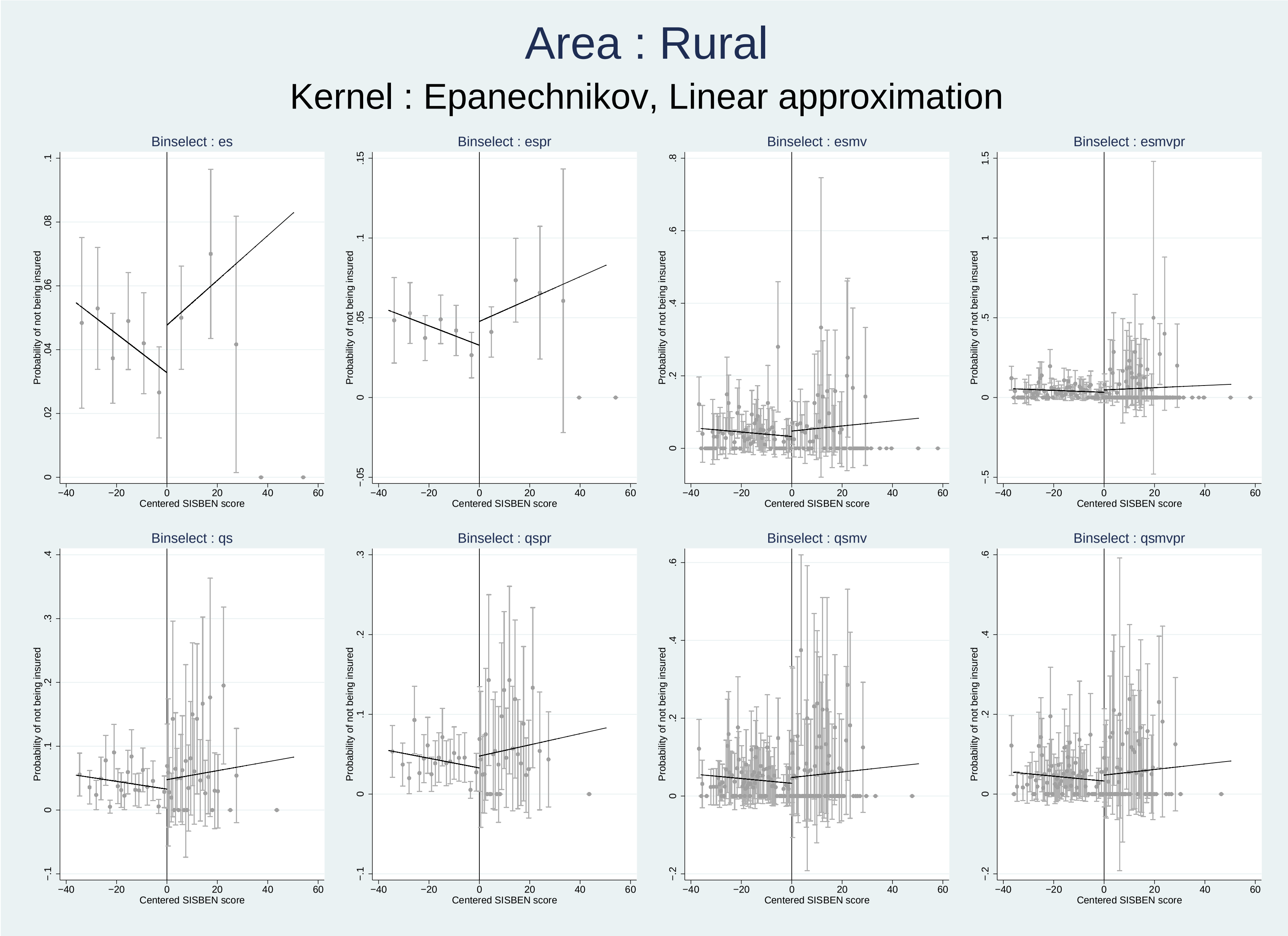
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the subsidized health insurance status. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Epanechnikov kernel using a linear approximation for the model around the threshold. 4./ Bin Selection Method: All methods available in STATA

Appendix 2E: Discontinuity identification: Contributive Regime. Different Bandwidth Select Methods (V)



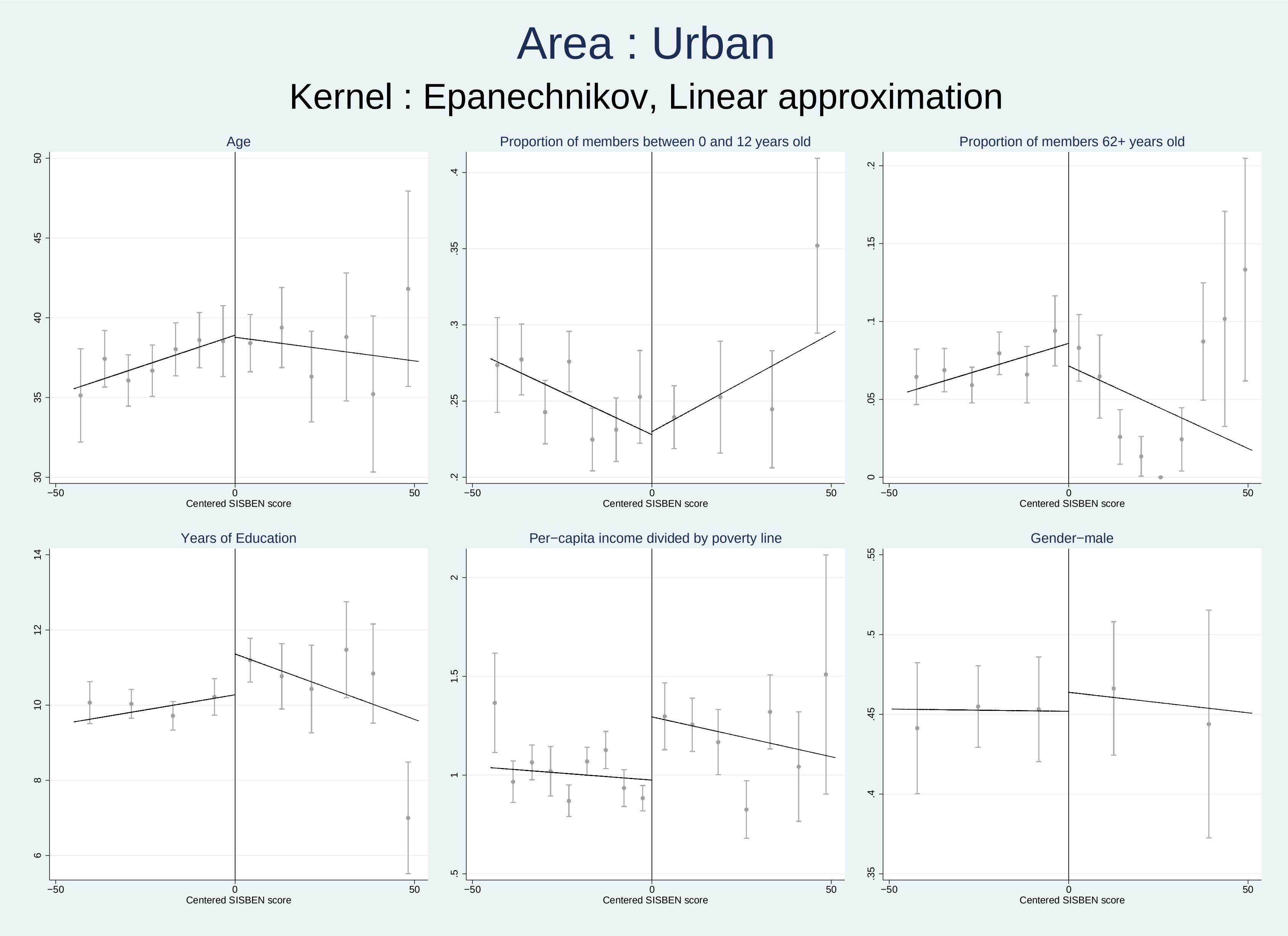
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the contributive health insurance status. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Epanechnikov kernel using a linear approximation for the model around the threshold. 4./ Bin Selection Method: All methods available in STATA

Appendix 2F: Discontinuity identification: Uninsured. Different Bandwidth Select Methods (VI)



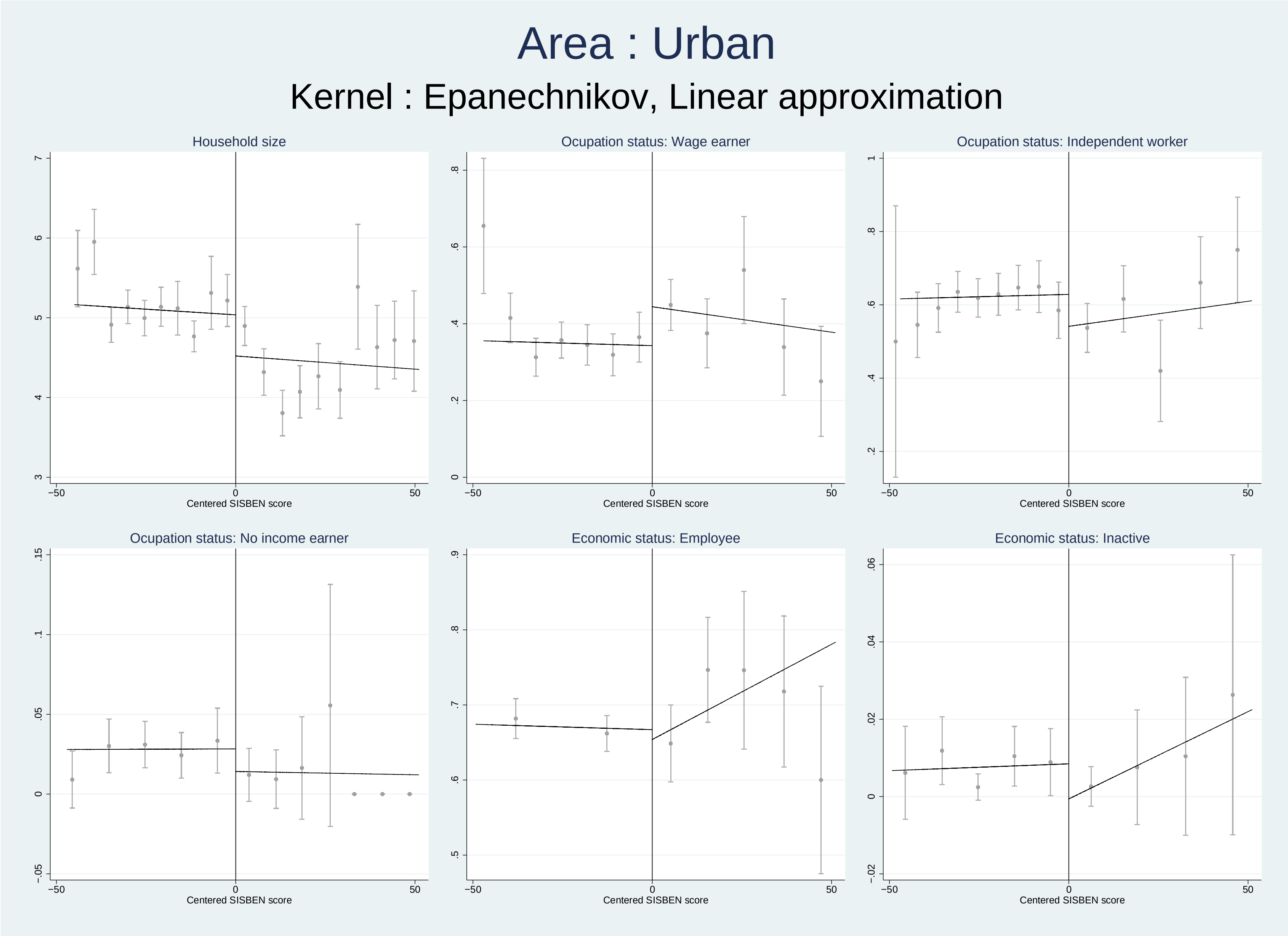
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the uninsured individuals. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Epanechnikov kernel using a linear approximation for the model around the threshold. 4./ Bin Selection Method: All methods available in STATA

Appendix 3A: Discontinuity identification for exogenous variables (I)



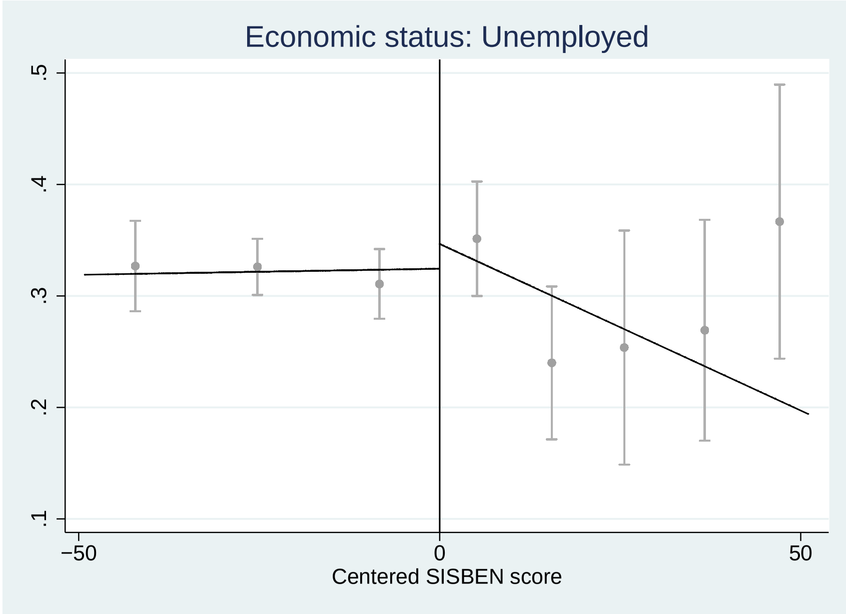
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the exogenous variables of the individuals included in the models. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: All methods available in STATA

Appendix 3B: Discontinuity identification for exogenous variables (II)



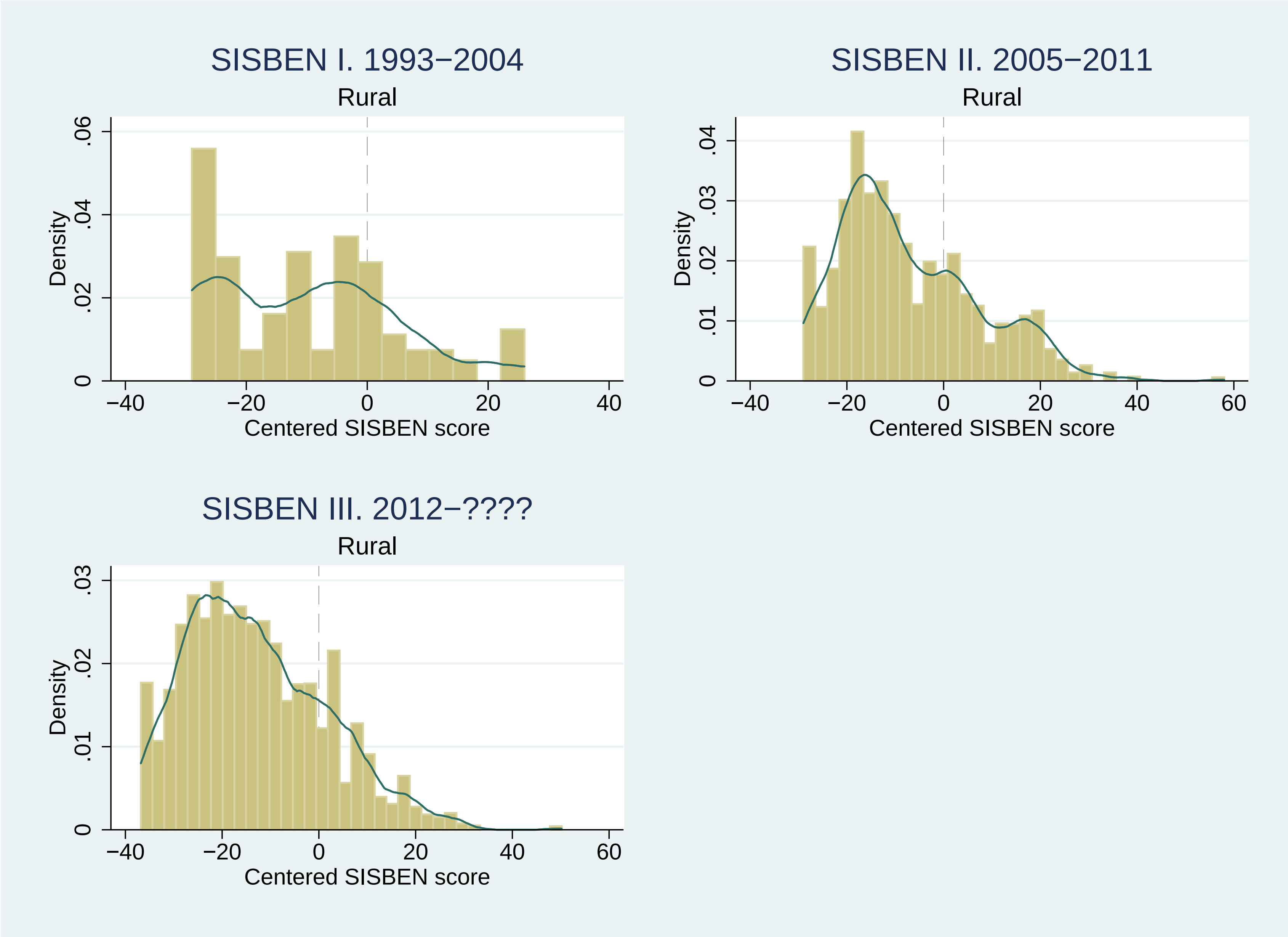
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the exogenous variables of the individuals included in the models. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: All methods available in STATA

Appendix 3C: Discontinuity identification for exogenous variables (III)



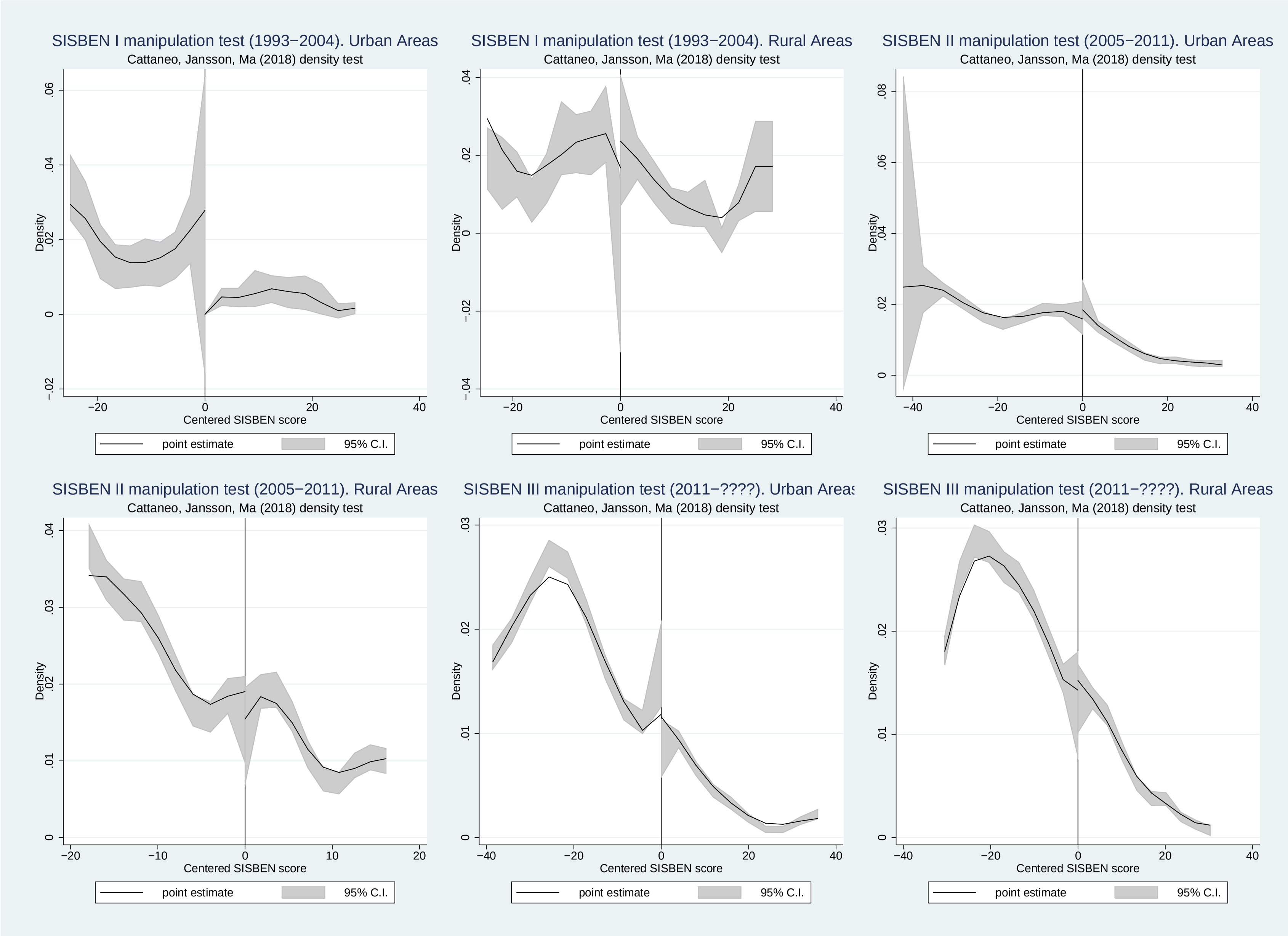
Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the exogenous variables of the individuals included in the models. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: All methods available in STATA

Appendix 4A: Rural SISBEN scores in each phase. Colombian Longitudinal Survey by Universidad de los Andes



Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Histogram and Kernel density estimates of the SISBEN score by household. 2./ SISBEN I, II and III separately and centered around zero. 3./ Restricting the sample only to households with individuals in the econometric models. 4./ Epanechnikov kernel

Appendix 4B: SISBEN manipulation test. Colombian Longitudinal Survey by Universidad de los Andes



Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Cattaneo, Jansson and Ma (2018) test. 2./ Density estimates for SISBEN I, II and III separately by Ruran/Urban area and centered around zero. 3./ Restricting the sample only to households with individuals in the econometric models.

Appendix 5A: Bonferroni correction test (I)

Variable	Benjamini and Liu for Conv. P-value	Benjamini and Liu for Rob. P-value	Number of Rejected null hipothesys Conv.	Number of Rejected null hipothesys Rob.
Thrombosis	0.0170	0.0170	0	0
Hearth attack	0.0170	0.0170	0	0
Hearth condition	0.0170	0.0170	0	0
Hypertension (Yes)	0.0170	0.0170	0	0
Hypertension (Only during pregnancy)	0.0170	0.0170	0	0
Asthma	0.0170	0.0170	0	0
Emphysema	0.0170	0.0170	0	0
Diabetes (Yes)	0.0170	0.0170	0	0
Diabetes (No)	0.0170	0.0170	0	0
Diabetes (Only during pregnancy)	0.0170	0.0170	0	0
Ulcer	0.0170	0.0170	0	0
Epilepsy	0.0500	0.0170	0	0
Cancer	0.0170	0.0170	0	0
Tuberculosis	0.0500	0.0170	0	0
Movement difficulties	0.0170	0.0170	0	0
Showering difficulties	0.0170	0.0170	0	0
Learning disability	0.0170	0.0170	0	0
Blindness	0.0170	0.0382	0	1
Illness	0.0170	0.0170	0	0
Accident	0.0170	0.0170	0	0
Dentistry	0.0170	0.0170	0	0
Surgery	0.0170	0.0170	0	0
Pregnancy	0.0170	0.0170	0	0
Fruit: Less than once a week	0.0170	0.0170	0	0
Fruit: Once a week	0.0170	0.0170	0	0
Fruit: 2-4 times a week	0.0170	0.0170	0	0
Fruit: 5-6 times a week	0.0170	0.0170	0	0
Fruit: Daily	0.0170	0.0170	0	0
Fruit: More than once a day	0.0170	0.0170	0	0

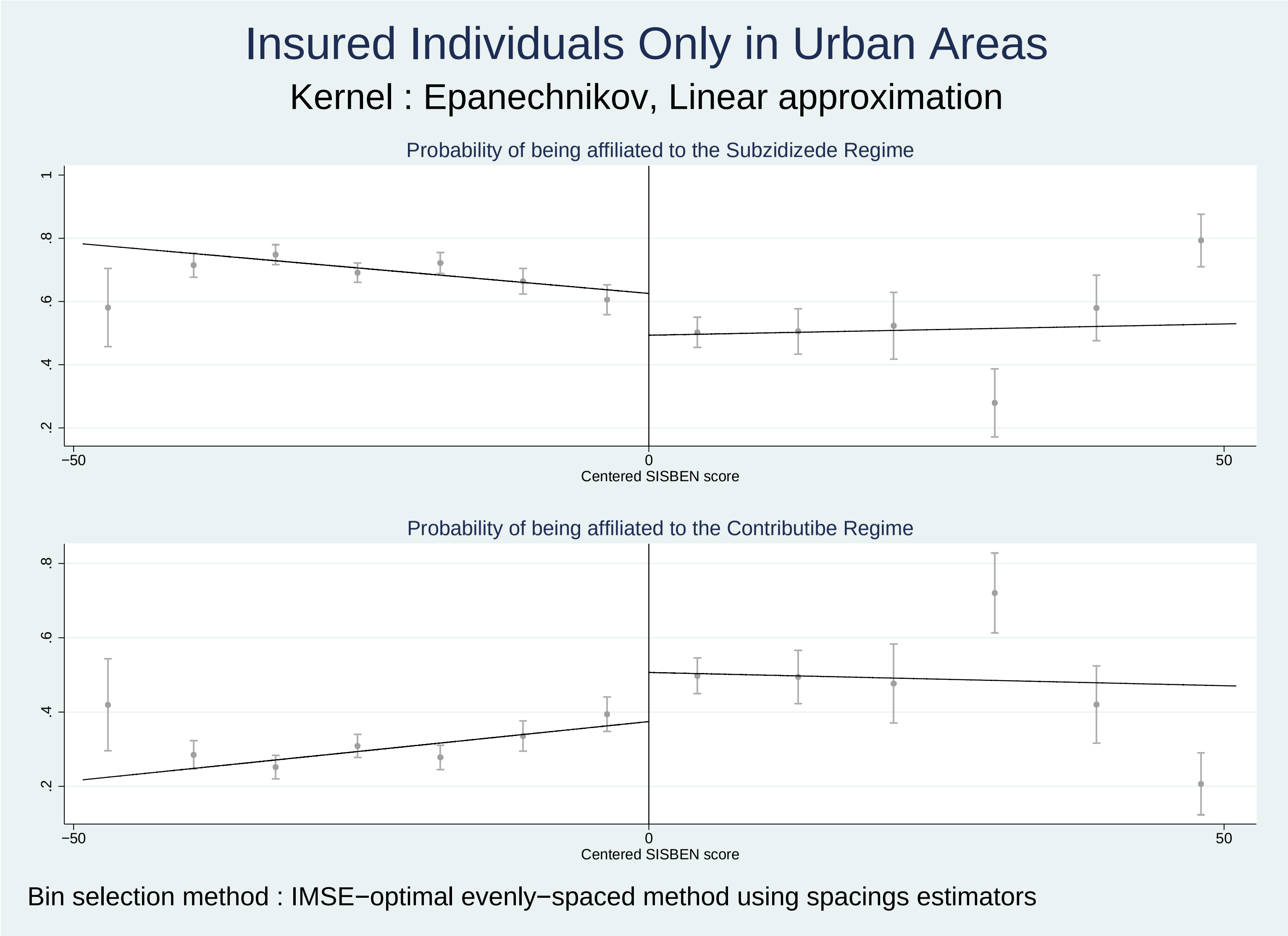
Source: Author's estimations using CLSA for the years 2010, 2013 and 2016.

Appendix 5B: Bonferroni correction test (II)

Variable	Benjamini and Liu for Conv. P-value	Benjamini and Liu for Rob. P-value	Number of Rejected null hipothesys Conv.	Number of Rejected null hipothesys Rob.
Vegetables: Less than once a week	0.0170	0.0170	0	0
Vegetables: 2-4 times a week	0.0170	0.0170	0	0
Vegetables: 5-6 times a week	0.0170	0.0170	0	0
Vegetables: Daily	0.0170	0.0170	0	0
Vegetables: More than once a day	0.0170	0.0170	0	0
Fried food: Less than once a month	0.0170	0.0170	0	0
Fried food: Once a month	0.0170	0.0170	0	0
Fried food: 2-3 times a month	0.0170	0.0170	0	0
Fried food: Twice a week	0.0170	0.0170	0	0
Fried food: 3-4 times a week	0.0170	0.0170	0	0
Fried food: 5-6 times a week	0.0170	0.0170	0	0
Fried food: Daily	0.0170	0.0170	0	0
Fried food: Twice a day	0.0170	0.0170	0	0
Fried food: Three or more a day	0.0170	0.0170	0	0
Moderate physical activity: Yes, in the last 7 days	0.0170	0.0170	0	0
Moderate physical activity: Number of days, in the last 7 days	0.0170	0.0170	0	0
Moderate physical activity: Minutes per day	0.0170	0.0170	0	0
Strong physical activity: Yes, in the last 7 days	0.0170	0.0170	0	0
Strong physical activity: Number of days, in the last 7 days	0.0170	0.0170	0	0
Strong physical activity: Minutes per day	0.0170	0.0170	0	0
Hospitalized in the last 12 months	0.0170	0.0170	0	0
Doctor	0.0170	0.0170	0	0
Dentist	0.0170	0.0170	0	0
Optometrist	0.0170	0.0170	0	0
Alternative Medicine	0.0170	0.0170	0	0
Contraceptive	0.0170	0.0170	0	0
Other	0.0170	0.0170	0	0
Household total monthly expenditure	0.0170	0.0170	0	0
Household food monthly expenditure	0.0170	0.0170	0	0
gasto_salud_anual	0.0170	0.0170	0	0

Source: Author's estimations using CLSA for the years 2010, 2013 and 2016.

Appendix 6: Discontinuity identification: Insured population



Source: Author's calculations using CLSA for the years 2010, 2013 and 2016. Note: 1./ Regression discontinuity estimation of the SISBEN score on the health insurance status. 2./ SISBEN I, II and III pooled and centered around zero. 3./ Restricting the sample only to individuals in the econometric models. 4./ Epanechnikov kernel using a linear approximation for the model around the threshold. 5./ Bin Selection Method: MSE optimal

Appendix 7A. Fuzzy Regression Discontinuity results: Health Status for Insured population

	Subsidized Regime		Contributive Regime		
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Conventional	Robust	Conventional	Robust	Observations
CIRCULATORY SYSTEM DISEASE - 2013 and 2016					
Thrombosis	5.960	59.43	0.280	0.195	1708
	(113.552)	(128.501)	(0.347)	(0.385)	1708
Hearth attack	0.0266	0.0292	-0.00910	-0.0192	1708
	(0.092)	(0.104)	(0.100)	(0.114)	1708
Hearth condition	0.425	0.239	-0.415	-0.284	1708
	(0.410)	(0.494)	(0.363)	(0.424)	1708
Hypertension (Yes)	2.970	-1.279	-1.013	-0.714	1708
	(5.292)	(6.455)	(0.958)	(1.090)	1708
Hypertension (Only during pregnancy)	0.291	0.639	-0.274	-0.601	1708
	(0.398)	(0.495)	(0.365)	(0.458)	1708
CHRONIC DISEASE - 2013 and 2016					
Asthma	-1.176	0.0445	1.185	-0.0596	1708
	(1.619)	(1.977)	(1.643)	(2.007)	1708
Emphysema	0.0407	0.0616	-0.0407	-0.0616	1708
	(0.124)	(0.144)	(0.124)	(0.143)	1708
Diabetes (Yes)	2.157	-3.267	-0.370	-0.283	1708
	(5.993)	(7.432)	(0.584)	(0.668)	1708
Diabetes (No)	-0.548	-0.478	0.497	0.451	1708
	(0.645)	(0.730)	(0.615)	(0.684)	1708
Diabetes (Only during pregnancy)	-0.467	-1.302	1.064	3.995	1708
	(0.947)	(1.150)	(4.000)	(4.669)	1708
Ulcer	0.0405	-0.102	-0.0474	0.130	1708
	(0.548)	(0.644)	(0.624)	(0.737)	1708
Epilepsy	0	0.00170	0	-0.00170	1708
	(0.000)	(0.004)	(0.000)	(0.004)	1708
Cancer	0.0736	0.0526	-0.0696	-0.0503	1708
	(0.073)	(0.083)	(0.072)	(0.081)	1708
INFECTIOUS DISEASES - 2013 and 2016					
Tuberculosis	-0.0101	-0.00560	0.00780	0.00460	1708
	(0.011)	(0.015)	(0.009)	(0.012)	1708
ACTIVITIES OF DAILY LIVING (ADL) - 2010, 2013 and 2016					
Movement difficulties	0.00340	-0.000100	-0.000800	0.00140	3945
	(0.061)	(0.073)	(0.055)	(0.065)	3945
HEALTH EVENT - 2010, 2013 and 2016					
Illness	-0.772	0.0314	0.192	0.0321	3845
	(0.840)	(0.991)	(0.543)	(0.615)	3845
Accident	-0.0879	-0.0784	0.0894	0.0789	3845
	(0.108)	(0.126)	(0.110)	(0.128)	3845
Dentistry	-0.443	-0.257	0.486	0.224	3845
	(0.363)	(0.426)	(0.412)	(0.493)	3845
Surgery	0	0.00190	0	-0.00200	3845
	(0.000)	(0.002)	(0.000)	(0.002)	3845
Pregnancy	-0.166	0.141	0.186	-0.279	1382
	(0.226)	(0.306)	(0.268)	(0.390)	1382

Source: Author's estimations using CLSA for the years 2013 and 2016 or 2010, 2013 and 2016 when indicated.
Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

Appendix 7B. Fuzzy Regression Discontinuity results: Behavioral distortions for Insured population

	Subsidized Regime		Contributive Regime		
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Conventional	Robust	Conventional	Robust	Observations
FRUIT CONSUMPTION - 2013 and 2016					
Less than once a week	0.0536	0.117	-0.0471	-0.110	1708
	(0.227)	(0.263)	(0.223)	(0.257)	1708
Once a week	0.192	0.0542	-0.192	-0.0519	1708
	(0.455)	(0.551)	(0.456)	(0.553)	1708
2-4 times a week	-0.0122	0.313	-0.417	-0.424	1708
	(0.606)	(0.719)	(0.458)	(0.505)	1708
5-6 times a week	0.417	0.108	-0.447	-0.0890	1708
	(0.582)	(0.693)	(0.615)	(0.738)	1708
Daily	-0.957	-0.928	1.307	0.954	1708
	(0.720)	(0.775)	(1.167)	(1.426)	1708
More than once a day	2.794	-2.416	-0.645	-0.617	1708
	(7.928)	(9.756)	(0.781)	(0.866)	1708
VEGETABLE CONSUMPTION - 2013 and 2016					
Less than once a week	0.327	0.271	-0.301	-0.270	1708
	(0.341)	(0.396)	(0.293)	(0.331)	1708
2-4 times a week	4.499	27.95	0.263	0.134	1708
	(30.289)	(36.761)	(0.961)	(1.099)	1708
5-6 times a week	-0.400	-0.284	1.273	-0.929	1708
	(0.722)	(0.860)	(2.918)	(3.650)	1708
Daily	-14.44	-163.3	-0.0226	-0.0665	1708
	(185.760)	(224.167)	(0.856)	(0.963)	1708
More than once a day	-1.441	-3.911	-4.624	10.28	1708
	(2.892)	(3.508)	(24.559)	(28.487)	1708
FRIED FOOD CONSUMPTION - 2013					
Less than once a month	0.501	0.0936	-1.476	2.257	1481
	(0.942)	(1.091)	(7.364)	(8.710)	1481
Once a month	-0.616	-0.521	0.642	0.512	1481
	(0.613)	(0.691)	(0.661)	(0.762)	1481
2-3 times a month	-0.716	-0.593	0.590	0.551	1481
	(0.663)	(0.758)	(0.463)	(0.505)	1481
Twice a week	2.013	1.287	-1.552	-1.438	1481
	(2.365)	(2.789)	(1.240)	(1.342)	1481
3-4 times a week	0.142	0.0172	-0.174	-0.0221	1481
	(0.862)	(0.998)	(0.889)	(1.036)	1481
5-6 times a week	0.158	0.0827	-0.163	-0.0804	1481
	(0.395)	(0.458)	(0.408)	(0.474)	1481
Daily	-1.332	-0.816	1.658	0.499	1481
	(1.167)	(1.342)	(1.622)	(1.969)	1481
Twice a day	-0.261	-0.147	0.259	0.141	1481
	(0.358)	(0.433)	(0.364)	(0.442)	1481
Three or more a day	0.0868	0.168	-0.0889	-0.178	1481
	(0.113)	(0.131)	(0.117)	(0.136)	1481
PHYSICAL ACTIVITY - 2013 and 2016					
Moderate: Yes in the last 7 days	-0.194	-0.166	0.201	0.163	1708
	(0.299)	(0.353)	(0.325)	(0.390)	1708
Moderate: Number of days in the last 7 days	-0.0795	0.511	0.194	-0.367	1708
	(2.612)	(3.222)	(2.245)	(2.725)	1708
Moderate: Minutes per day	-18.09	-9.552	17.30	10.58	1704
	(33.493)	(40.737)	(32.599)	(38.790)	1704
Strong: Yes in the last 7 days	0.183	0.163	-0.212	-0.164	1708
	(0.173)	(0.196)	(0.224)	(0.264)	1708
Strong: Number of days in the last 7 days	0.132	0.164	-0.205	-0.228	1708
	(0.142)	(0.170)	(0.242)	(0.318)	1708
Strong: Minutes per day	60.52	25.46	-70.60	-19.57	1702
	(58.740)	(71.105)	(66.784)	(82.963)	1702

Source: Author's estimations using CLSA for the years 2013 or 2013 and 2016 when indicated.
Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

Appendix 7C. Fuzzy Regression Discontinuity results: Medical Care Use for Insured population

	Subsidized Regime		Contributive Regime		
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Conventional	Robust	Conventional	Robust	Observations
MEDICAL CARE USE - 2010, 2013 and 2016					
Hospitalized in the last 12 months	-0.431	-0.338	0.597	0.239	3845
	(0.405)	(0.458)	(0.610)	(0.740)	3845
Doctor	0.289	0.175	-0.373	-0.219	3845
	(0.527)	(0.577)	(0.572)	(0.632)	3845
Dentist	-0.586	-0.566	0.721	0.604	3845
	(0.548)	(0.592)	(0.783)	(0.902)	3845
Optometrist	0.243	0.0305	-0.346	0.238	3845
	(0.601)	(0.665)	(0.829)	(0.973)	3845
Alternative Medicine	1.260	5.455	-4.163	-43.02	3845
	(4.789)	(5.684)	(48.764)	(57.104)	3845
Contraceptive	0.729	1.936	-0.479	-1.134	2484
	(1.937)	(2.306)	(1.082)	(1.312)	2484
Other	-0.0932	-0.0600	0.0969	0.0629	2851
	(0.263)	(0.307)	(0.264)	(0.307)	2851

Source: Author's estimations using CLSA for the years 2010, 2013 and 2016 when indicated.
Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.

Appendix 7C: Fuzzy Regression Discontinuity results: Risk Protection and Consumption Smoothing for Insured population

	Subsidized Regime		Contributive Regime		
	(1)	(2)	(3)	(4)	(2)
VARIABLES	Conventional	Robust	Conventional	Robust	Observations
RISK PROTECTION AND CONSUMPTION SMOOTHING - 2010, 2013 and 2016					
Log Household total monthly expenditure	0.151	0.0762	-0.144	-0.0675	4800
	(0.261)	(0.282)	(0.262)	(0.284)	4800
Log Household food monthly expenditure	0.0264	-0.00140	-0.00830	0.0184	4800
	(0.340)	(0.368)	(0.345)	(0.375)	4800
Log Household health annual expenditure	-4.262	-2.498	3.426	2.019	611
	(2.631)	(3.252)	(2.458)	(2.942)	611

Source: Author's estimations using CLSA for the years 2010, 2013 and 2016 when indicated.
Note: *Significant at 10%, ** at 5%, *** at 1%,. Standard errors in parentheses. Odd columns present the results for the conventional estimates with conventional standard errors, while even columns do it for the bias-corrected estimator coupled with the robust variance estimators. The order of the approximation of the polynomial for estimating the conventional coefficient is 1, while the order of the approximation for estimating the bias-correction is 2. Triangular kernel is used for the estimations and a data-driven bandwidth selection method.